APPENDIX

Nonuniform Space-Time Codes for Layered Source Coding

Erik G. Larsson and Wing-Hin Wong
Dept. of Electrical and Computer Engineering
University of Florida
P.O. Box 116130
Gainesville, FL 32611, USA
Email: larsson@dsp.ufl.edu
Phone: 352-392-2633
Fax: 352-392-0044

Abstract

We propose new space-time codes tailored to point-to-multipoint, or broadcast, communications using layered source coding. Our codes can be encoded (and decoded) differentially, and they are based entirely on phase-shift keying. We develop analytical design criteria for the codes, and we discuss the design of optimal and suboptimal receiver structures. We also discuss the relation between our new codes and a differentially encoded Alamouti code. Numerical examples illustrate the performance of our new codes.

Index Terms

space-time coding, diversity, maximum-likelihood, differential encoding, joint estimation/detection, nonuniform constellations, phase-shift keying

I. Introduction

During the last decade, wireless systems with multiple transmit and receive antennas have been studied extensively, and the performance of such systems has been proven to be extremely promising. The propagation channel associated with a system with multiple transmit and receive antennas is sometimes called a *multiple-input multiple-output* (MIMO) channel, and the associated coding and signal processing is referred to as *space-time coding*. Using proper space-time coding, it is possible to use the degrees of freedom of the MIMO channel both to increase the throughput and to counteract fading. For this reason, MIMO technology is believed to become a major cornerstone in many future wireless communication systems.

Loosely speaking, communication links can be classified into two categories: point-to-point links and point-to-multipoint links. In the former case, there is exactly one transmitter and one receiver which communicate with each other at a given time, whereas for a point-to-multipoint link, a transmitted message is aimed at multiple different recipients simultaneously. A cellular communication system with mobile users is an example of a point-to-point communication system, whereas a radio/TV broadcast is an example of a point-to-multipoint link (therefore point-to-multipoint links are sometimes referred to as broadcast channels).

Point-to-multipoint links are becoming increasingly important. For instance, the introduction of Digital Audio Broadcast (DAB) and High-Definition Television (HDTV) has pioneered a whole new field of digital broadcasting applications. As a further example, it is widely believed that much of the next generation's wireless networking will be based on so-called ad-hoc networks, where it may be necessary for multiple units to listen to one message at the same time. For some early and fundamental research on broadcast channels the reader is referred to [1] and [2].

There are two major differences between point-to-point and point-to-multipoint communication links. First, a point-to-point connection can be optimized for a given transmitter-receiver pair. For instance, a cellular system usually employs power control techniques that adjust the transmitted power to minimize the

power consumption, reduce the amount of co-channel interference, and at the same time ensure that the received signal strength exceeds a certain threshold. Second, in contrast to a point-to-point communication link, for a broadcast transmission *all receivers have different capabilities to decode the transmitted message.* This is so because the different receivers experience in general very different radio link qualities (for instance, due to varying propagation conditions). Moreover, since quality can usually be traded for cost, the receivers themselves may have different *inherent* abilities to decode the transmitted message.

The fact that distinct receivers have different capabilities of decoding a message suggests that the transmitted signal should consist of several components which are of different importance for the reconstruction of the message (and therefore have an inherently different vulnerability to transmission errors). This idea has lead to the concept of layered source coding (see, e.g., [3], [4], [5], [6]), which is now a mature technique employed in many multimedia standards. For instance, the image coding standard JPEG-2000 and the video coding standard MPEG-4 use what is sometimes referred to as "fine granularity scalability," which enables a gradual tradeoff between the error-free data throughput and the quality of the reconstructed image or video sequence [7]. Such progressive source coding methods are already in use in many Internet applications where data rate can be traded for quality, and they are expected to play an instrumental role for the next generation of wireless standards to provide ubiquitous access both to the Internet, and to diverse sources of streaming video and audio.

The topic of this paper is space-time coding for broadcast channels when layered source coding is used. Previous works on the topic include [8], which discussed "beamforming" techniques (hence requiring feedback of channel state information) for a multiuser system. Reference [9] proposed a layered space-time coding scheme assuming different receivers may have different numbers of receive antennas, and that a receiver with more receive antennas can decode more layers of messages. In this paper, we suggest a new layered space-time coding scheme that does not require any channel knowledge at the transmitter, and which is constructed starting from a criterion that attempts to minimize the error rate. Our new codes also satisfy two other important design goals. First, they can be encoded (and detected) differentially. Second, they are entirely based on phase-shift keying (PSK) modulation and consequently the transmitted signal has constant envelope at all times. The codes proposed in this paper can therefore be seen as a multidimensional extension of a technique in [10].

II. Nonuniform SISO PSK Codes

With layered source coding, for instance as used in JPEG-2000 and MPEG-4, more capable receivers can achieve a higher data rate by decoding the messages contained in all layers while the less capable receivers can only decode the message in the bottom layer. Thus the "most important" information should be conveyed via the bottom layer while "less important" information can be transmitted via higher layer(s).

An example of a modulation scheme designed for layered source coding is nonuniform phase-shift-keying modulation (PSK), which was first introduced by Pursley and Shea [10] (see also [11] and [12]). The idea is intuitively appealing. Starting with a standard uniform PSK constellation, which contains data for the "*basic*" layer, a nonuniform PSK code constellation is obtained by adding a usually small additional phase shift α to each original constellation point, which contains the information for the "*additional*" layer. This is illustrated in Figure 1, when the *basic* message is conveyed via a QPSK constellation. A capable receiver with high signal-to-noise-ratio (SNR) can distinguish among all eight constellation points which means that both the *basic* and the *additional* layer messages can be decoded, whereas to a less capable receiver the constellation may appear like a blurred QPSK constellation and hence it may only be able to distinguish between the different "large" phase shifts; thus only the *basic* message can be detected accurately. By such a construction the error rates associated with the basic and the additional message are different. Consequently, the additional message can carry information that is of less importance for the reconstruction of the transmitted message than the basic message is. The error probability of the additional message can be easily adjusted by choosing different values of the additional phase shift α. Obviously, a larger α results in a better performance for the additional message at the cost of a deterioration of the performance for the

basic message. The main virtue of this nonuniform PSK encoding scheme is that the signaling envelope is constant while multi-layered messages can be transmitted with different and adjustable error performance.

III. Space-Time Coding

Consider a MIMO system with n_t transmit antennas, n_r receive antennas, and assume for simplicity that the propagation channel is linear, time-invariant and frequency flat. Let $\boldsymbol{H}$ be an $n_r \times n_t$ matrix whose (m,n)th element contains the channel gain between transmit antenna n and receive antenna m, and suppose that a code matrix $\boldsymbol{X}$ of dimension $n_t \times N$ is taken from a matrix constellation $\mathcal{X}$ and transmitted by sending its N columns via the n_t antennas during N time epochs. If the signal received during the same N time intervals is arranged in an $n_r \times N$ matrix $\boldsymbol{Y}$, then:

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E} \tag{1}$$

where $\boldsymbol{E}$ is an $n_r \times N$ matrix of noise. Throughout this paper, we shall make the somewhat standard assumption that the elements of $\boldsymbol{H}$ and $\boldsymbol{E}$ are independent and zero-mean complex Gaussian with variances ρ^2 and σ^2, respectively; hence the channel is Rayleigh fading.

Given $\boldsymbol{Y}$, detection of $\boldsymbol{X}$ in a maximum-likelihood (ML) sense amounts to minimizing the Euclidian distance

$$\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\| \tag{2}$$

with respect to $\boldsymbol{X} \in \mathcal{X}$, as well as $\boldsymbol{H}$ (unless it is known). The average (over $\boldsymbol{H}$) probability that the ML detector mistakes a transmitted code matrix $\boldsymbol{X}_0$ for an incorrect (and different) code matrix $\boldsymbol{X} \neq \boldsymbol{X}_0$ can be bounded by [13], [14], [15]:

$$E\left[P(\boldsymbol{X}_0 \to \boldsymbol{X})\right]\Big|_{\text{coherent}} \leq \left|(\boldsymbol{X}-\boldsymbol{X}_0)(\boldsymbol{X}-\boldsymbol{X}_0)^H\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \tag{3}$$

and

$$E\left[P(\boldsymbol{X}_0 \to \boldsymbol{X})\right]\Big|_{\text{noncoherent}} \leq \left|\boldsymbol{X}_0 \Pi^{\perp}_{\boldsymbol{X}^H} \boldsymbol{X}_0^H\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \tag{4}$$

for the two cases that the channel $\boldsymbol{H}$ is known and unknown to the receiver respectively. Here $|\cdot|$ denotes the determinant of a matrix and $(\cdot)^H$ stands for the complex conjugate transpose.

Note that if $\boldsymbol{H}$ is unknown, we can obtain a detection rule that depends only on $\boldsymbol{X}$. Consider the problem of minimizing (2):

$$\min_{\boldsymbol{X}\in\mathcal{X},\boldsymbol{H}} \ \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|^2 \tag{5}$$

Treating $\boldsymbol{H}$ as a deterministic unknown, we can minimize this cost function with respect to $\boldsymbol{H}$. The result is (see, e.g., [16] and [15]):

$$\boldsymbol{H} = \boldsymbol{Y}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1} \tag{6}$$

Inserting (6) into (5) yields:

$$\min_{\boldsymbol{X}\in\mathcal{X}} \ \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1}\boldsymbol{X}\|^2 \Longleftrightarrow \min_{\boldsymbol{X}\in\mathcal{X}} \ \|\boldsymbol{Y}\Pi^{\perp}_{\boldsymbol{X}^H}\|^2 \tag{7}$$

where

$$\begin{aligned} \Pi_{\boldsymbol{X}} &= \boldsymbol{X}(\boldsymbol{X}^H\boldsymbol{X})^{-1}\boldsymbol{X}^H \\ \Pi^{\perp}_{\boldsymbol{X}} &= \boldsymbol{I} - \Pi_{\boldsymbol{X}} \end{aligned} \tag{8}$$

stand for the projections onto the range space of $\boldsymbol{X}$ and its orthogonal complement, and $\boldsymbol{I}$ is the identity matrix.

IV. Differential Modulation for MIMO Systems

Space-time modulation matrices that can be encoded *differentially* are often of special interest since such codes can easily be demodulated noncoherently. Differential codes with uniform error probabilities have been studied by many authors (see, e.g., [17], [18], [19], [20], [21] for some prominent examples) and some of them can be seen as an extension of differential PSK to MIMO systems. In general, if t is the time index and if $\{\boldsymbol{U}(t)\}$ is a sequence of (square) information-bearing matrices, then differential encoding obtains the transmitted code matrix $\boldsymbol{X}(t)$ at time t via:

$$\boldsymbol{X}(t) = \boldsymbol{X}(t-1)\boldsymbol{U}(t) \tag{9}$$

where $\boldsymbol{X}(t-1)$ is the matrix transmitted at time $t-1$. Such an encoding usually is only meaningful under certain circumstances, for instance if $\boldsymbol{U}(t)$ is unitary (in which case $\boldsymbol{X}(t)$ becomes unitary for all t as well, given a unitary "initial matrix" $\boldsymbol{X}(0)$). By considering two (or more) received matrices $\boldsymbol{Y}(t)$ and $\boldsymbol{Y}(t-1)$ simultaneously, noncoherent detection is possible. For example, if we concatenate the matrices received at time $t-1$ and t, and assume that the channel $\boldsymbol{H}$ remains constant over these two time intervals (to within practical accuracy), we can write (using (1) and (9)):

$$\begin{aligned}
&\big[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\big] \\
=&\big[\boldsymbol{H}\boldsymbol{X}(t-1) \quad \boldsymbol{H}\boldsymbol{X}(t)\big] + \big[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t)\big] \\
=&\big[\boldsymbol{H}\boldsymbol{X}(t-1) \quad \boldsymbol{H}\boldsymbol{X}(t-1)\boldsymbol{U}(t)\big] + \big[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t)\big] \\
=&\boldsymbol{H}\boldsymbol{X}(t-1)\cdot\big[\boldsymbol{I} \quad \boldsymbol{U}(t)\big] + \big[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t)\big]
\end{aligned} \tag{10}$$

where $\boldsymbol{H}\boldsymbol{X}(t-1)$ can be seen as an unknown "effective" channel matrix and $\big[\boldsymbol{I} \quad \boldsymbol{U}(t)\big]$ is an effective code matrix. Provided that $\big[\boldsymbol{I} \quad \boldsymbol{U}(t)\big]$ is such that the corresponding determinant in (4) is nonzero, noncoherent demodulation is possible. By applying the decision rule (7) to the data model (10), the detection rule for differential demodulation becomes:

$$\begin{aligned}
&\min_{\boldsymbol{U}(t)\in\mathcal{X}} \left\| \big[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\big]\cdot\Pi^{\perp}_{[\boldsymbol{I}\ \boldsymbol{U}(t)]^H} \right\|^2 \\
\Longleftrightarrow &\min_{\boldsymbol{U}(t)\in\mathcal{X}} \mathrm{Tr}\left\{ \big[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\big]\cdot\Pi^{\perp}_{[\boldsymbol{I}\ \boldsymbol{U}(t)]^H}\cdot\begin{bmatrix}\boldsymbol{Y}^H(t-1)\\ \boldsymbol{Y}^H(t)\end{bmatrix}\right\} \\
\Longleftrightarrow &\min_{\boldsymbol{U}(t)\in\mathcal{X}} \mathrm{Tr}\left\{ \big[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\big]\cdot\begin{bmatrix}\boldsymbol{I} & -\boldsymbol{U}(t)\\ -\boldsymbol{U}^H(t) & \boldsymbol{I}\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{Y}^H(t-1)\\ \boldsymbol{Y}^H(t)\end{bmatrix}\right\} \\
\Longleftrightarrow &\max_{\boldsymbol{U}(t)\in\mathcal{X}} \mathrm{Re}\left\{ \mathrm{Tr}\left\{\boldsymbol{U}(t)\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1)\right\}\right\}
\end{aligned} \tag{11}$$

where $\mathrm{Tr}\{\cdot\}$ stands for the trace of a matrix and where we used the fact that the information-bearing matrix $\boldsymbol{U}(t)$ is unitary. The associated average error probability, i.e., the probability that $\boldsymbol{U}^{(0)}$ is wrongly detected as $\boldsymbol{U}$ for this differentiation detection scheme can be bounded by using (4):

$$\begin{aligned}
E\left[P(\boldsymbol{U}^{(0)}\to\boldsymbol{U})\right] &\le \left|\big[\boldsymbol{I} \quad \boldsymbol{U}^{(0)}\big]\cdot\Pi^{\perp}_{[\boldsymbol{I}\ \boldsymbol{U}]^H}\cdot\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{U}^{(0)H}\end{bmatrix}\right|^{-n_r}\cdot\left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \\
&= \left|\boldsymbol{I}-\frac{1}{2}(\boldsymbol{U}\boldsymbol{U}^{(0)H}+\boldsymbol{U}^{(0)}\boldsymbol{U}^H)\right|^{-n_r}\cdot\left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t}
\end{aligned} \tag{12}$$

For simplicity hereafter the time index t of $\boldsymbol{U}(t)$ is omitted whenever no confusion can occur.

V. Nonuniform Space-Time Codes

Inspired by the nonuniform PSK codes discussed in Section II, our new nonuniform space-time codes are based on differential encoding of the product of a unitary code matrix $\boldsymbol{U}_b \in \mathcal{X}_b$ associated with a basic message, and another unitary code matrix $\boldsymbol{U}_a \in \mathcal{X}_a$ corresponding to an additional message. Thus the transmitted code matrix at time t is given by (cf. (9)):

$$\boldsymbol{X}(t) = \boldsymbol{X}(t-1)\boldsymbol{U}_b(t)\boldsymbol{U}_a(t) \tag{13}$$

Using (12) along with the fact that $\boldsymbol{U}_b\boldsymbol{U}_a$ is unitary, we find that the average probability that a transmitted message pair $(\boldsymbol{U}_b^0,\boldsymbol{U}_a^0)$ is mistaken for another pair $(\boldsymbol{U}_b,\boldsymbol{U}_a)$ can be bounded by:

$$\begin{aligned} &E\left[P(\boldsymbol{U}_b^0,\boldsymbol{U}_a^0 \to \boldsymbol{U}_b,\boldsymbol{U}_a)\right] \\ &\quad \le \left|\boldsymbol{I} - \frac{1}{2}\left(\boldsymbol{U}_b\boldsymbol{U}_a\boldsymbol{U}_a^{0H}\boldsymbol{U}_b^{0H} + \boldsymbol{U}_b^0\boldsymbol{U}_a^0\boldsymbol{U}_a^H\boldsymbol{U}_b^H\right)\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \end{aligned} \tag{14}$$

A bound such as (12) or (14) was used in, for instance, [17] for the design of (uniform) differential space-time codes. However, although it may be thought of as a feasible approach, an attempt to minimize this bound in the context of *nonuniform* space-time modulation may *not* produce the desired result since the target error rates for $\boldsymbol{U}_b$ and $\boldsymbol{U}_a$ are different.

A. Design Criteria

Since the additional message can be decoded only at high SNR, the matrices $\{\boldsymbol{U}_a\}$ associated with the additional message should approximately be close to the identity matrix. Inspired by this observation, we can first consider the design of $\{\boldsymbol{U}_b\}$, treating the presence of $\boldsymbol{U}_a$ as an unmodelled noise-like disturbance term. Doing so, we can approximately bound the error probability for $\boldsymbol{U}_b$ alone (assuming differential detection) by:

$$E\left[P(\boldsymbol{U}_b^0 \to \boldsymbol{U}_b)\right] \lesssim \left|\boldsymbol{I} - \frac{1}{2}\left(\boldsymbol{U}_b^{(0)}\boldsymbol{U}_b^H + \boldsymbol{U}_b\boldsymbol{U}_b^{(0)H}\right)\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2+\alpha^2}\right)^{-n_r n_t} \tag{15}$$

where α^2 is a factor that incorporates the noise-like effect of the presence of $\boldsymbol{U}_a$. Although the "bound" (15) is somewhat heuristic (and probably neither tight nor very accurate), we believe that it may serve a purpose as a meaningful design criterion for $\{\boldsymbol{U}_b\}$.

Next, for the design of $\{\boldsymbol{U}_a\}$ we proceed as follows. Suppose that the SNR is in a region such that $\boldsymbol{U}_b$ can be reliably decoded. For the design of $\{\boldsymbol{U}_a\}$ this should be a reasonable assumption, since if it is not true then decoding of $\boldsymbol{U}_a$ is probably of less interest anyway. Assuming that $\boldsymbol{U}_b$ is known, the demodulation of $\boldsymbol{U}_a$ is essentially another noncoherent detection problem. To obtain a criteria for the design of the constellation $\{\boldsymbol{U}_a\}$, we want to form an error bound on $\boldsymbol{U}_a$ and average it over all possible basic messages $\{\boldsymbol{U}_b\}$. Using the bound (4) for noncoherent detection appears to yield criteria that are very difficult to use. As a suboptimal approach we used instead the bound (3) for coherent detection. Doing so results in the following criterion:

$$\begin{aligned} E\left[P(\boldsymbol{U}_a^0 \to \boldsymbol{U}_a)\right] &\le \frac{1}{|\mathcal{X}_b|}\sum_{\boldsymbol{U}_b\in\mathcal{X}_b} \left|(\boldsymbol{U}_b\boldsymbol{U}_a^0 - \boldsymbol{U}_b\boldsymbol{U}_a)(\boldsymbol{U}_b\boldsymbol{U}_a^0 - \boldsymbol{U}_b\boldsymbol{U}_a)^H\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \\ &= \left|(\boldsymbol{U}_a^0 - \boldsymbol{U}_a)(\boldsymbol{U}_a^0 - \boldsymbol{U}_a)^H\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \end{aligned} \tag{16}$$

where in the last step $\boldsymbol{U}_b$ disappears since it is unitary. In (16), $|\cdot|$ denotes the number of elements of the set.

B. Union Bound on the Error Probability

By the union bound, the probability for the basic message to be in error (averaged over all possible pairs of additional messages) can be bounded by:

$$\begin{aligned} &E\left[P(\text{basic message in error})\right] \\ \le &\frac{1}{|\mathcal{X}_b|} \sum_{\substack{(\boldsymbol{U}_b^{(k)}, \boldsymbol{U}_b^{(n)}) \in \mathcal{X}_b \\ k \neq n}} \frac{1}{|\mathcal{X}_a|^2} \sum_{(\boldsymbol{U}_a^{(r)}, \boldsymbol{U}_a^{(s)}) \in \mathcal{X}_a} \\ &\left| \boldsymbol{I} - \frac{1}{2}\left(\boldsymbol{U}_b^{(k)} \boldsymbol{U}_a^{(r)} \boldsymbol{U}_a^{(s)H} \boldsymbol{U}_b^{(n)H} + \boldsymbol{U}_b^{(n)} \boldsymbol{U}_a^{(s)} \boldsymbol{U}_a^{(r)H} \boldsymbol{U}_b^{(k)H} \right) \right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2} \right)^{-n_r n_t} \end{aligned} \tag{17}$$

Likewise, the error rate for the additional message (averaged over all possible pairs of basic messages), can be bounded by a similar expression:

$$\begin{aligned} &E\left[P(\text{additional message in error})\right] \\ \le &\frac{1}{|\mathcal{X}_a|} \sum_{\substack{(\boldsymbol{U}_a^{(r)}, \boldsymbol{U}_a^{(s)}) \in \mathcal{X}_a \\ r \neq s}} \frac{1}{|\mathcal{X}_b|^2} \sum_{(\boldsymbol{U}_b^{(k)}, \boldsymbol{U}_b^{(n)}) \in \mathcal{X}_b} \\ &\left| \boldsymbol{I} - \frac{1}{2}\left(\boldsymbol{U}_b^{(k)} \boldsymbol{U}_a^{(r)} \boldsymbol{U}_a^{(s)H} \boldsymbol{U}_b^{(n)H} + \boldsymbol{U}_b^{(n)} \boldsymbol{U}_a^{(s)} \boldsymbol{U}_a^{(r)H} \boldsymbol{U}_b^{(k)H} \right) \right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2} \right)^{-n_r n_t} \end{aligned} \tag{18}$$

These bounds will be useful for performance optimization.

VI. Design Examples

A. $R_b = 1$, $R_a = 1$ Code for 2 TX Antennas

We first construct a code where the rate for the basic message is $R_b = 1$ bit/sec/Hz and the rate for the additional message is $R_a = 1$ bit/sec/Hz as well. We take the basic message $\boldsymbol{U}_b$ taken from the following set:

$$\mathcal{X}_b = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\} \tag{19}$$

The constellation in (19), which is uniform and possesses certain optimality properties, is due to Hughes [17] and was called "BPSK" therein. Based on the design rules in Section V-A, we have handcrafted the following constellation of matrices for the additional message $\boldsymbol{U}_a$:

$$\mathcal{X}_a = \left\{ \begin{bmatrix} e^{i\pi\lambda} & 0 \\ 0 & e^{i\pi\gamma} \end{bmatrix}, \begin{bmatrix} e^{i\pi\gamma} & 0 \\ 0 & e^{i\pi\lambda} \end{bmatrix}, \begin{bmatrix} e^{-i\pi\lambda} & 0 \\ 0 & e^{-i\pi\gamma} \end{bmatrix}, \begin{bmatrix} e^{-i\pi\gamma} & 0 \\ 0 & e^{-i\pi\lambda} \end{bmatrix} \right\} \tag{20}$$

where (λ, γ) are design parameters (to be discussed below).

The constellation may appear to be somewhat arbitrary but it possesses some nice features. It is symmetric and uniform which means that the error probability performance should be the same for all constellation points. Also similarly to the nonuniform PSK encoding for a single channel (cf. Section II), multiplication with the matrices in (20) can be interpreted as adding a small phase shift to the elements of the basic message $\boldsymbol{U}_b$.

If we take, somewhat arbitrarily:

$$\boldsymbol{X}(0) = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \tag{21}$$

it follows that

$$\boldsymbol{X}^H(t)\boldsymbol{X}(t) = 2\boldsymbol{I} \tag{22}$$

for all t and hence the differential encoding of the code is meaningful. Furthermore, we can verify that all elements of $\boldsymbol{X}$ have always unit magnitude:

$$\left|X_{k,l}(t)\right| = 1 \tag{23}$$

and hence all transmitted symbols are obtained by constant envelope modulation, which was one of our design goals.

The error performance (of both the basic and the additional message) will depend on λ and γ, and typically the error performance of the additional message can be traded against that of the basic message. The values (λ,γ) must be chosen carefully. For instance, at first glance intuition would perhaps suggest us to take $\lambda > 0$ and $\gamma = 0$, but we can show that such a "simple" choice leads to a code that "works" but that does *not* provide maximal diversity.

In our experiments, we used the union bound (17) and (18) to optimize over (λ, γ). In particular, for a given "tolerable" loss in performance for the basic message, we can find the pair (λ, γ) that minimizes (18). The result is shown in Figure 2. For example, if we can accept a degradation of 1.5 dB for the basic message, then $\lambda = 0.2$ and $\gamma = 0.035$ are optimal. The optimization over (λ, γ) is further illustrated in Figure 3, where we show how the performance associated with the basic and additional messages varies with (λ, γ).

Figure 4 shows the empirical bit-error-rate (BER), obtained via Monte-Carlo simulation, for the code described above using $\lambda = 0.2$, $\gamma = 0.035$ and ML decoding. The solid lines ("—") show the performance of differential nonuniform BPSK for a conventional system with $n_t = 1$ transmit antenna and a single receive antenna (this is essentially a special case of [10]). The dashed lines ("- -") show the performance for a system with $n_t = 2$ (and a single receive antenna) using the new code presented above. For the curves without marks, only a basic message is transmitted. The curves with marks show the performance when both a basic and an additional message are transmitted: the curves marked with "o" show the BER for the basic message, and the curves with "×" show the BER for the additional message. Clearly, the transmit diversity system outperforms the conventional one – observe, in particular, the different slopes of the BER curves both for the basic and for the additional message. The simulation also confirms that the transmission of an additional message incurs a small performance degradation for the basic message.

B. $R_b = 2$, $R_a = 1$ Code

To obtain a code with a higher information rate for the basic message, we next take $\boldsymbol{U}_b$ from the following algebraic group of 16 matrices (which is due to [17] as well) generated by:

$$\mathcal{X}_b = \left\langle \begin{bmatrix} e^{j\pi/4} & 0 \\ 0 & e^{-j\pi/4} \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle \tag{24}$$

The notation in (24) means that all possible matrices can be generated by choosing any arbitrary integers M and N in the following expression:

$$\begin{bmatrix} e^{j\pi/4} & 0 \\ 0 & e^{-j\pi/4} \end{bmatrix}^M \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^N \tag{25}$$

The constellation used for the additional message is chosen to be the same as (20), so we now have a system with rate $R_b = 2$ for the basic message and rate $R_a = 1$ for the additional message. The corresponding simulated BER is shown in Figure 5.

VII. Suboptimal Detector for Nonuniform D-STBC

To detect both the basic and the additional messages in a joint ML fashion we have to search through all $|\mathcal{X}_b| \cdot |\mathcal{X}_a|$ possible combinations to find the combination of $\boldsymbol{U}_b$ and $\boldsymbol{U}_a$ which maximize the cost function in (11). This may be computationally burdensome if the constellation sizes of both layers are large. A possible remedy to this problem is to use a suboptimal detection rule described as follows:

Step 1. Detect the basic message only using (7), neglecting the presence of the additional message:

$$\begin{aligned} \hat{\boldsymbol{U}}_b(t) &= \min_{\boldsymbol{U}_b(t)\in\mathcal{X}_b} \left\| \left[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\right] \cdot \Pi^{\perp}_{[I\ \boldsymbol{U}_b(t)]^H} \right\|^2 \\ &\Longleftrightarrow \max_{\boldsymbol{U}_b(t)\in\mathcal{X}_b} \mathrm{Re}\left\{\mathrm{Tr}\left\{\boldsymbol{U}_b(t)\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1)\right\}\right\} \end{aligned} \tag{26}$$

Step 2. Detect the additional message assuming the decision taken in *Step 1* is correct:

$$\begin{aligned} &\min_{\boldsymbol{U}_a\in\mathcal{X}_a} \left\| \left[\boldsymbol{Y}(t-1)\cdot\hat{\boldsymbol{U}}_b(t) \quad \boldsymbol{Y}(t)\right] \cdot \Pi^{\perp}_{[I\ \boldsymbol{U}_a]^H} \right\|^2 \\ &\Longleftrightarrow \max_{\boldsymbol{U}_a(t)\in\mathcal{X}_a} \mathrm{Re}\left\{\mathrm{Tr}\left\{\boldsymbol{U}_a(t)\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1)\hat{\boldsymbol{U}}_b(t)\right\}\right\} \end{aligned} \tag{27}$$

Let us elaborate more on why this suboptimal detection scheme works. Since $\boldsymbol{U}_a$ is close to the identity matrix, when compared to the basic message $\boldsymbol{U}_b$ it can be treated as small unmodelled noise; hence the detection rule in *Step 1* can still be able to detect the basic message with only a small performance degradation. As the error performance of the basic message is typically superior to that of the additional message, we can assume the detected basic message is correct when we are dealing with the detection of the additional message. Thus the observation model for the detection of $\boldsymbol{U}_a(t)$ becomes:

$$\begin{aligned} &\left[\boldsymbol{Y}(t-1)\cdot\hat{\boldsymbol{U}}_b(t) \quad \boldsymbol{Y}(t)\right] \\ &\quad = \boldsymbol{H}\boldsymbol{X}(t-1)\cdot\left[\hat{\boldsymbol{U}}_b(t) \quad \hat{\boldsymbol{U}}_b(t)\boldsymbol{U}_a(t)\right] + \left[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t)\right] \\ &\quad = \boldsymbol{H}\boldsymbol{X}(t-1)\hat{\boldsymbol{U}}_b(t)\cdot\left[\boldsymbol{I} \quad \boldsymbol{U}_a(t)\right] + \left[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t)\right] \end{aligned} \tag{28}$$

Treating $\boldsymbol{H}\boldsymbol{X}(t-1)\hat{\boldsymbol{U}}_b(t)$ as an unknown "effective" channel as in (10) and using (7) and (11) yields the detection rule in *Step 2*.

By applying this suboptimal detection rule the computational complexity is reduced from $|\mathcal{X}_b| \cdot |\mathcal{X}_a|$ to $|\mathcal{X}_b| + |\mathcal{X}_a|$. For example, for the $R_b = 2$, $R_a = 1$ code, the computational time is reduced from $|\mathcal{X}_b| \cdot |\mathcal{X}_a| = 16 \cdot 4 = 64$ to $|\mathcal{X}_b| + |\mathcal{X}_a| = 16 + 4 = 20$. For this example the reduction is not very significant; yet this suboptimal detection scheme does provide more flexibility as opposed to the joint ML detection because if only the *basic* message is needed, the computational load will be further reduced to $|\mathcal{X}_b| = 16$ (for this example).

We expect that there is some performance loss associated with this suboptimal detection scheme since it does not exploit knowledge of the structure of the additional message in the *Step 1*. The Monte-Carlo simulation result in Figure 6 shows the performance of the suboptimal detector for the $R_b = 2$, $R_a = 1$ code compared to the performance of the joint ML detector. The detection of the basic message gives a small performance loss ($\approx$ 0.3 dB) compared to the joint ML detector, whereas the detection of the additional message is (to within practical accuracy) equal to that of the joint ML detector.

VIII. Comparison of Differential Group Codes and a Differential Alamouti Code

A. Differential Alamouti Code

The main advantage of the nonuniform space-time group codes introduced in Section VI is that the signal envelope is constant. However, also for this reason they cannot achieve the maximum possible performance

(in terms of BER) because they do not make use of any amplitude information in the signal. On the contrary, orthogonal space-time block codes (OSTBC) should provide better performance (BER) since they are already optimized by construction (see [22], [23] and [15]). For 2 TX antennas OSTBC reduces to Alamouti code. The construction of an Alamouti code matrix is as follows:

$$\boldsymbol{U}(t) = \frac{1}{\sqrt{2}}\begin{bmatrix} s_1(t) & s_2^*(t) \\ s_2(t) & -s_1^*(t) \end{bmatrix} \tag{29}$$

where s_1 and s_2 are the information bearing symbols and $(\cdot)^*$ stands for the complex conjugate. For the $R_b = 1$ system, s_1 and s_2 are taken from the BPSK set $\{-1, 1\}$, and for the $R_b = 2$ system, they are taken from the QPSK set $\{-1, -j, 1, j\}$. The $1/\sqrt{2}$ factor is included for normalization purposes so that the transmitted power per time interval equals one. This Alamouti matrix $\boldsymbol{U}(t)$ can then be encoded differentially by using (9) in Section IV. Note that the Alamouti code works together with the additional message $\boldsymbol{U}_a$ taken from (20) to form a nonuniform constellation, and differentially encoded by (13), but unfortunately the matrices in the so-obtained constellation do not have constant signal envelope.

Here we compare the performance of the group codes to that of the Alamouti code. The $R_b = 1$ group codes (19) in Section VI and $R_b = 1$ Alamouti codes are indeed equivalent, so there is no performance advantage for Alamouti codes for this case. The equivalence of both codes is easy to show as in this case the set of Alamouti matrices becomes:

$$\mathcal{X}_{\text{alamouti}} = \left\{ \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \right\} \tag{30}$$

If we pre-multiply the above set by $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ then we obtain the $R_b = 1$ group code in (19).

For the $R_b = 2$ codes, the Alamouti code does provide somewhat better BER performance than the group codes in (24) and (20). This is so since the Alamouti code uses both signal amplitute and phase to transmit information; thus it should be less susceptible to the phase disturbance caused by the additional message. Figure 7 shows the simulated performance difference between the $R_b = 2$, $R_a = 1$ nonuniform group codes in (24) and (20) and a $R_b = 2$ Alamouti code together with the additional message (20). For the basic message the Alamouti code outperforms the group code by about 1dB, and it is somewhat more robust when additional message is added. Meanwhile the performances of the detection of the additional message are equal both when the Alamouti and the group codes (20) are used as the basic message carrier.

B. Alamouti Code with Differentially Encoded Symbols

It may be argued that both the constant signaling envelope and maximum BER performance can be achieved at the same time by using symbol-wise differential encoding and forming the Alamouti matrix afterwards. However, unfortunately this encoding scheme does not work. This is demonstrated as follows.

Suppose that we employ symbol-wise differential encoding:

$$\begin{aligned} x_1(t) &= x_1(t-1) \cdot s_1(t) \\ x_2(t) &= x_2(t-1) \cdot s_2(t) \end{aligned} \tag{31}$$

and use these elements to form the Alamouti matrix:

$$\begin{aligned} \boldsymbol{X}(t-1) &= \begin{bmatrix} x_1(t-1) & x_2^*(t-1) \\ x_2(t-1) & -x_1^*(t-1) \end{bmatrix} \\ \boldsymbol{X}(t) &= \begin{bmatrix} x_1(t) & x_2^*(t) \\ x_2(t) & -x_1^*(t) \end{bmatrix} = \begin{bmatrix} x_1(t-1)s_1(t) & x_2^*(t-1)s_2^*(t) \\ x_2(t-1)s_2(t) & -x_1^*(t-1)s_1^*(t) \end{bmatrix} \end{aligned} \tag{32}$$

The above code matrix has the property that $|x_1(t)| = |x_2(t)| = |s_1(t)| = |s_2(t)| = 1$ for all t and hence the signal envelope is constant for both transmit antennas all the time. Also $\boldsymbol{X}(t)$ is unitary for all t since it is an Alamouti matrix. The observation model (10) becomes (cf. Section IV):

$$\begin{aligned}&\left[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\right]\\ =&\boldsymbol{H}\left[\boldsymbol{X}(t-1) \quad \boldsymbol{X}(t)\right] + \left[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t)\right]\\ =&\boldsymbol{H}\boldsymbol{X}(t-1)\left[\boldsymbol{I} \quad \boldsymbol{X}^H(t-1)\boldsymbol{X}(t)\right] + \left[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t)\right]\end{aligned} \tag{33}$$

Employing the noncoherent detection rule (11 to the above model (33) the following detection rule is obtained:

$$\begin{aligned}&\max_{s_1(t),s_2(t)\in\mathcal{S}} \mathrm{Re}\left\{ \mathrm{Tr}\left\{(\boldsymbol{X}^H(t-1)\boldsymbol{X}(t)\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1))\right\}\right\}\\ \Longleftrightarrow &\max_{s_1(t),s_2(t)\in\mathcal{S}}\\ &\mathrm{Re}\left\{ \mathrm{Tr}\left\{\begin{bmatrix} x_1^*(t-1) & x_2^*(t-1)\\ x_2(t-1) & -x_1(t-1)\end{bmatrix}\begin{bmatrix} x_1(t-1)\cdot s_1(t) & x_2^*(t-1)\cdot s_2^*(t)\\ x_2(t-1)\cdot s_2(t) & -x_1^*(t-1)\cdot s_1^*(t)\end{bmatrix}\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1)\right\}\right\}\\ \Longleftrightarrow &\max_{s_1(t),s_2(t)\in\mathcal{S}}\\ &\mathrm{Re}\quad \mathrm{Tr}\quad \begin{matrix} s_1(t)+s_2(t) & x_1^*(t-1)x_2^*(t-1)\cdot(s_2^*(t)-s_1^*(t))\\ x_1(t-1)x_2(t-1)\cdot(s_1(t)-s_2(t)) & s_1^*(t)+s_2^*(t)\end{matrix} \boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1)\end{aligned} \tag{34}$$

The above maximization problem (34) cannot be solved unless $x_1(t-1)$ and $x_2(t-1)$ are known. For example, consider the following two sets of values:

$$\{s_1(t), s_2(t), x_1(t-1)x_2(t-1)\} = \{1, -1, 1\} \tag{35}$$

and

$$\{s_1(t), s_2(t), x_1(t-1)x_2(t-1)\} = \{-1, 1, -1\} \tag{36}$$

These make the metric in (34) equal to:

$$\mathrm{Re}\left\{ \mathrm{Tr}\left\{\begin{bmatrix} 0 & -2\\ 2 & 0\end{bmatrix}\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1)\right\}\right\} \tag{37}$$

which implies that the sets $\{s_1, s_2\} = \{1, -1\}$ and $\{s_1, s_2\} = \{-1, 1\}$ are not identifiable if $x_1(t-1)$ and $x_2(t-1)$ are unknown.

IX. A Network Application Example

We consider employing our nonuniform space-time codes in a broadcasting telecommunication system. For simplicity we assume that the base station is located at the center of a circular cell. Define the outage radius to be the radius of the area such that a specified detection performance (in terms of BER) is achievable with a certain probability that we call the Quality of Service (QoS). For example, if we let R_0 denote the 99% outage radius then within the area of radius R_0 the message can be detected at a specified BER during 99% the of total time.

If nonuniform space-time codes are used to increase the data rate for the more capable receivers, it is clear that the outage radius for the transmission of the basic message under the same QoS will decrease to $R_b < R_0$. Meanwhile the outage radius of the transmission of the additional message now becomes R_a; of course its value depends on QoS for the additional message. For the simplicity of our analysis here let us assume the same QoS for both the basic and the additional message.

Denote the "tolerable" performance loss for the detection of the *basic* message as Δ dB. Assuming log-distance fading it is straightforward to show that:

$$\frac{R_b^2}{R_0^2} = 10^{-\frac{\Delta}{10}\cdot\frac{2}{n}} \tag{38}$$

$$\frac{R_a^2}{R_0^2} = 10^{-\frac{\Gamma}{10}\cdot\frac{2}{n}} \tag{39}$$

where Γ is the performance difference (in dB) between the detection of the *additional* message and that of the standard uniform constellation system. The parameter n is the path loss exponent for the radio link environment (i.e.: Power $\propto$ (distance)$^{-n}$; for details see [24]). The above equations are also valid in a log-normal fading model.

Figure 8 shows how R_b and R_a vary with Δ for the $R_b = 1$, $R_a = 1$ system, with the power attenuation factor of the radio link taken to be $n = 2$ and $n = 4$; these are typical values for free-space and downtown areas, respectively [24]. It can be seen that employing the nonuniform constellation does *not* increase the total throughput of the system (i.e., the sum of the data rate for all users) in the case $n = 2$. For example, if the outage area for the basic message is allowed to decrease to 0.8 times of that of the original cell area, the outage area of the additional message can only reach 0.15 times of that of the original cell area, which results in a decrease in overall system throughput. However for the case $n = 4$, the use of a nonuniform code increases the total throughput. This observation is not hard to understand; for $n = 4$ the area closely surrounding area a base station enjoys a relatively much higher SNR than the area close to the border of the cell and hence the outage area for the additional can be extended further. Note that the outage areas of the nonuniform layered code (19) and (20) (with $R_b = R_a = 1$) tend to approach that of the $R_b = 2$ code (24) when Δ increases.

X. Concluding Remarks

We have presented new nonuniform space-time codes that can be encoded and detected differentially, and that are based entirely on phase-shift keying. We also discussed analytical criteria for code construction and optimization, and we compared its performance with that of a scheme based on the Alamouti code. We also studied a suboptimal detector and its performance, which we found to be satisfactory (a loss within 0.3 dB compared to joint ML decoding). We also demonstrated how nonuniform space-time codes in a broadcasting telecommunication system can increase the total throughput.

It may be argued that using already established nonuniform constellations for single transmit antenna systems together with, for instance, known linear space-time codes should be a natural approach to the problem of designing nonuniform space-time constellations. However, there are at least two problems associated with such an approach. First, it may in general *not* be optimal, simply because we are optimizing over the class of space-time codes and the class of nonuniform single-antenna constellations *separately*, instead of optimizing over the class of nonuniform MIMO constellations. Second, we found it difficult to incorporate constraints (such as constant envelope after differential encoding), that are desired or required from a practical implementation point of view. Therefore, we believe that it may be advantageous to design nonuniform space-time constellations by approaching the problem from first principles.

References

[1] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, 1972.

[2] P. P. Bergmans and T. M. Cover, "Cooperative broadcasting," *IEEE Transactions on Information Theory*, vol. 20, pp. 317–324, May 1974.

[3] L.-F. Wei, "Coded modulation with unequal error probabilities," *IEEE Transactions on Communications*, vol. 41, pp. 1439–1449, Oct. 1993.

[4] A. R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Transactions on Information Theory*, vol. 39, pp. 1234–1248, July 1993.

[5] M. Sajadieh, F. R. Kschischang, and A. Leon-Garcia, "Modulation-assisted unequal error protection over the fading channel," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 900–908, Aug. 1998.

[6] K. Ramchandran, A. Ortega, K. M. Uz, and M. Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 6–23, 1993.
[7] W. Li, "Overview of fine granularity scalability in the MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 301–317, Mar. 2001.
[8] M. Memarzadeh, A. Sabharwal, and B. Aazhang, "Broadcast space-time coding," in *Proc. of the SPIE International Symposium on the Convergence of Information Theory and Communications (ITCOM)*, 2001.
[9] C.-H. Kuo, C.-S. Kim, R. Ku, and C.-C. J. Kuo, "Embedded space-time coding for wireless broadcast," in *Proc. of Visual Communication and Image Processing*, 2002. To appear.
[10] M. B. Pursley and J. M. Shea, "Nonuniform phase-shift key modulation for multimedia multicast transmission in mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 774–783, May 1999.
[11] M. B. Pursley and J. M. Shea, "Multimedia multicast wireless communications with phase-shift key modulation and convolutional coding," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1999–2010, 1999.
[12] M. B. Pursley and J. M. Shea, "Adaptive nonuniform phase-shift key modulation for multimedia traffic in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1394–1407, 2000.
[13] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, Mar. 1998.
[14] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, pp. 543–564, Mar. 2000.
[15] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003. To appear.
[16] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, UK: Prentice Hall International, 1989.
[17] B. L. Hughes, "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, pp. 2567–2578, Nov. 2000.
[18] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1169–1174, July 2000.
[19] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Transactions on Communications*, vol. 48, pp. 2041–2052, Dec. 2000.
[20] G. Ganesan and P. Stoica, "Differential modulation using space-time block codes," *IEEE Communications Letters*, vol. 9, pp. 57–60, Feb. 2002.
[21] X. G. Xia, "Differential en/decoding orthogonal space-time block codes with APSK signals," *IEEE Communications Letters*, vol. 6, pp. 150–152, Apr. 2002.
[22] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, Oct. 1998.
[23] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.
[24] T. Rappaport, *Wireless Communications: Principles and Practice*. Prentice-Hall, 2002.

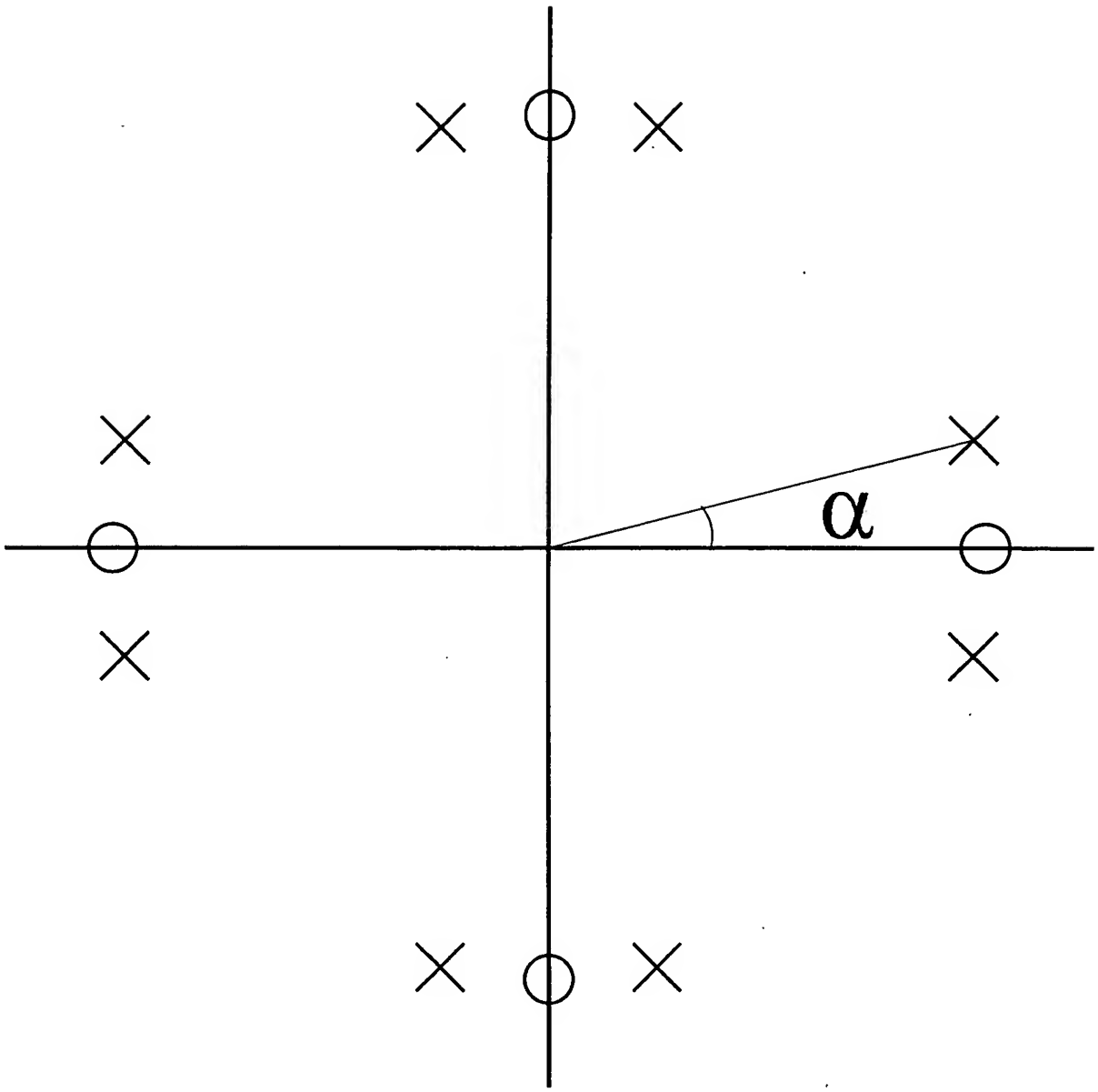

Fig. 1. A nonuniform 8-PSK constellation obtained from a standard "uniform" QPSK constellation by splitting each original constellation point "o" into a pair of new points "×".

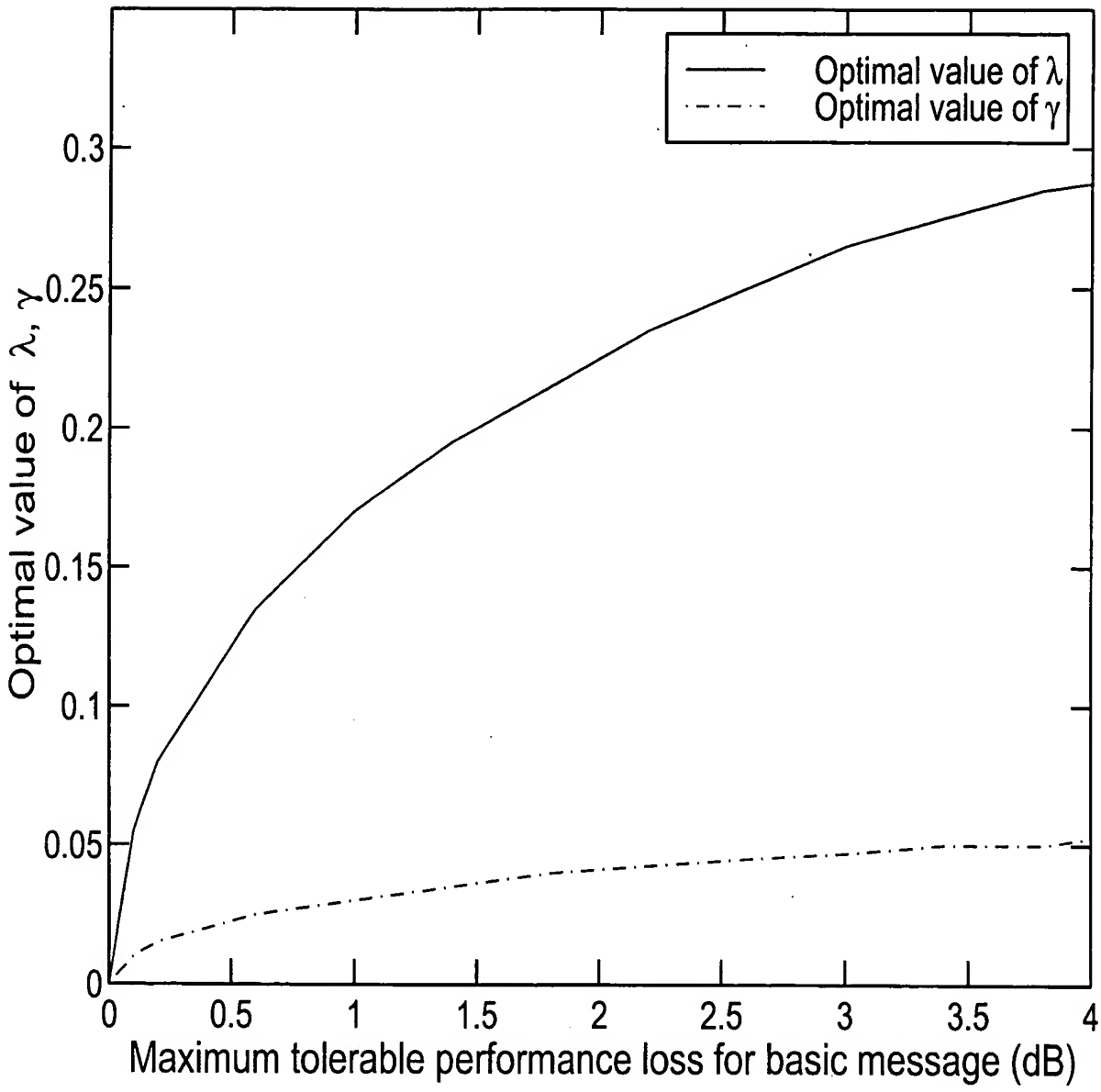

Fig. 2. Optimal values of (λ, γ) for the $R_b = 1$, $R_a = 1$ code, for a given acceptable performance degradation of the basic message. The results are obtained via minimization of the union bound (17) and a corresponding expression for the error rate of the additional message.

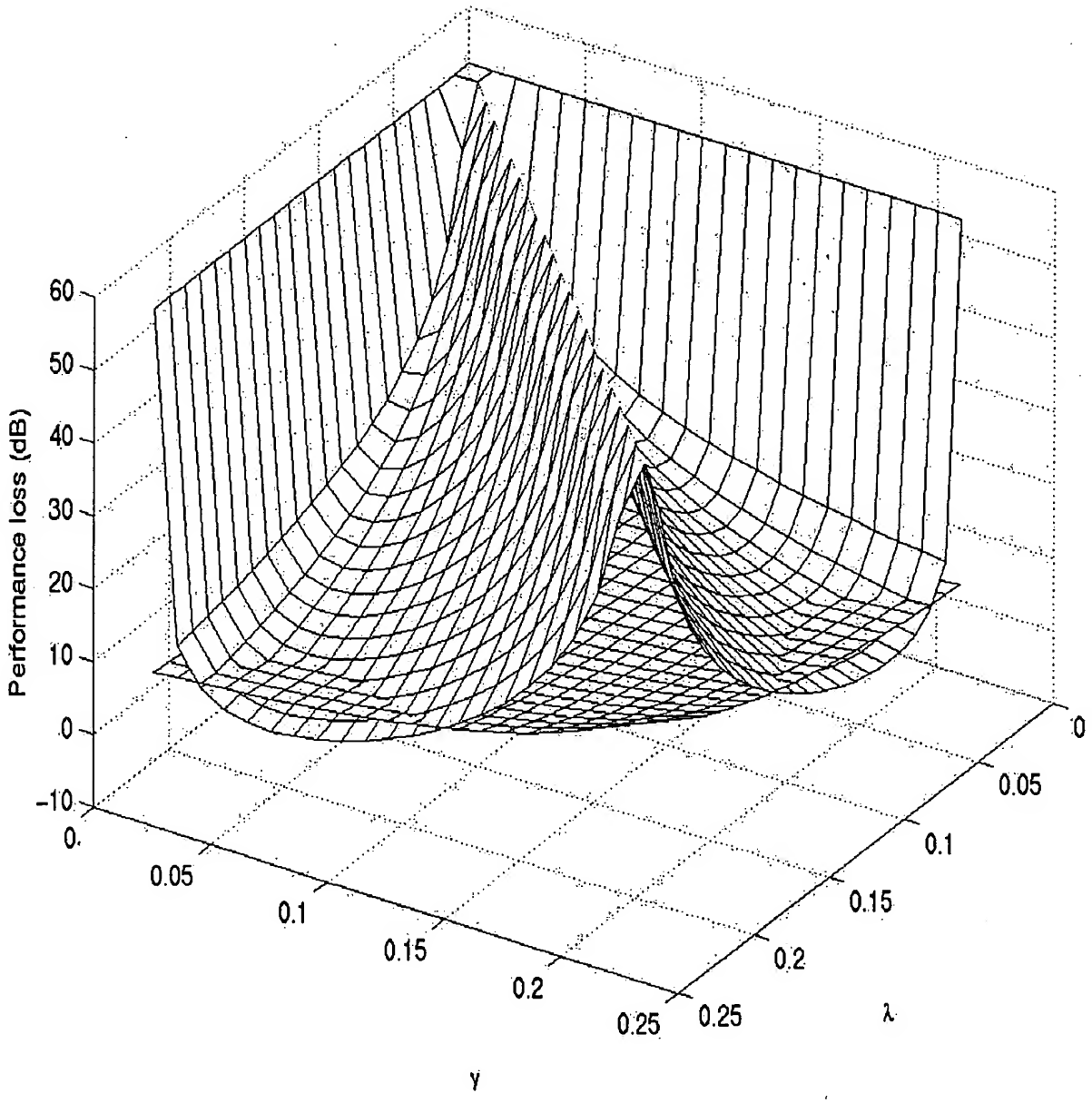

Fig. 3. Performance degradation for the basic message and performance gain for the additional message for the $R_b = 1$, $R_a = 1$ code. The results are obtained via (17), along with a corresponding expression for the additional message. The curves for the basic message are normalized relative to the "undisturbed case" ($\lambda = \gamma = 0$) and the curves for the additional message are normalized relative to $\lambda = 0.2$, $\gamma = 0.035$.

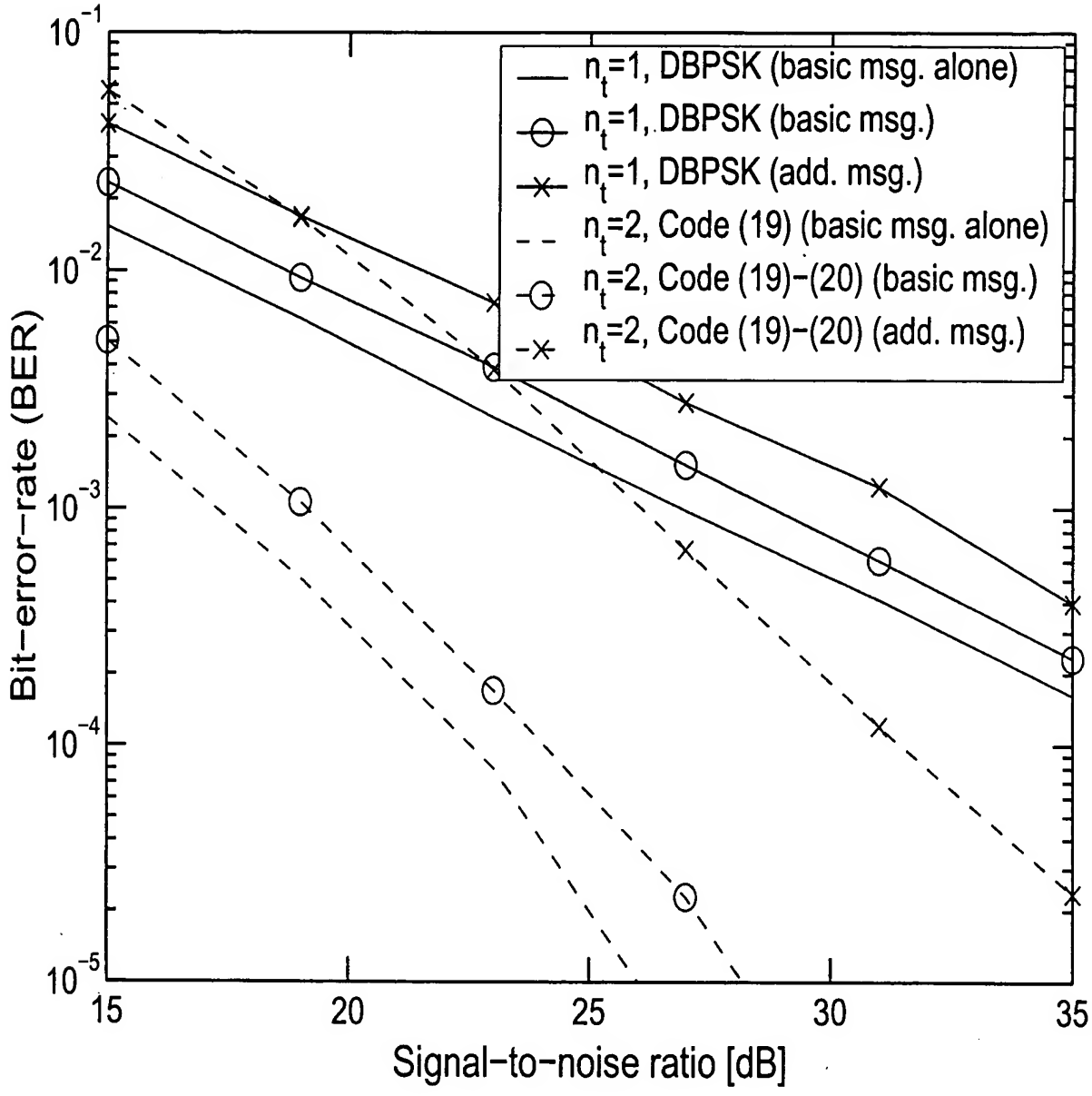

Fig. 4. Empirical BER for the $R_b = 1$, $R_a = 1$ code with $\lambda = 0.2$, $\gamma = 0.035$.

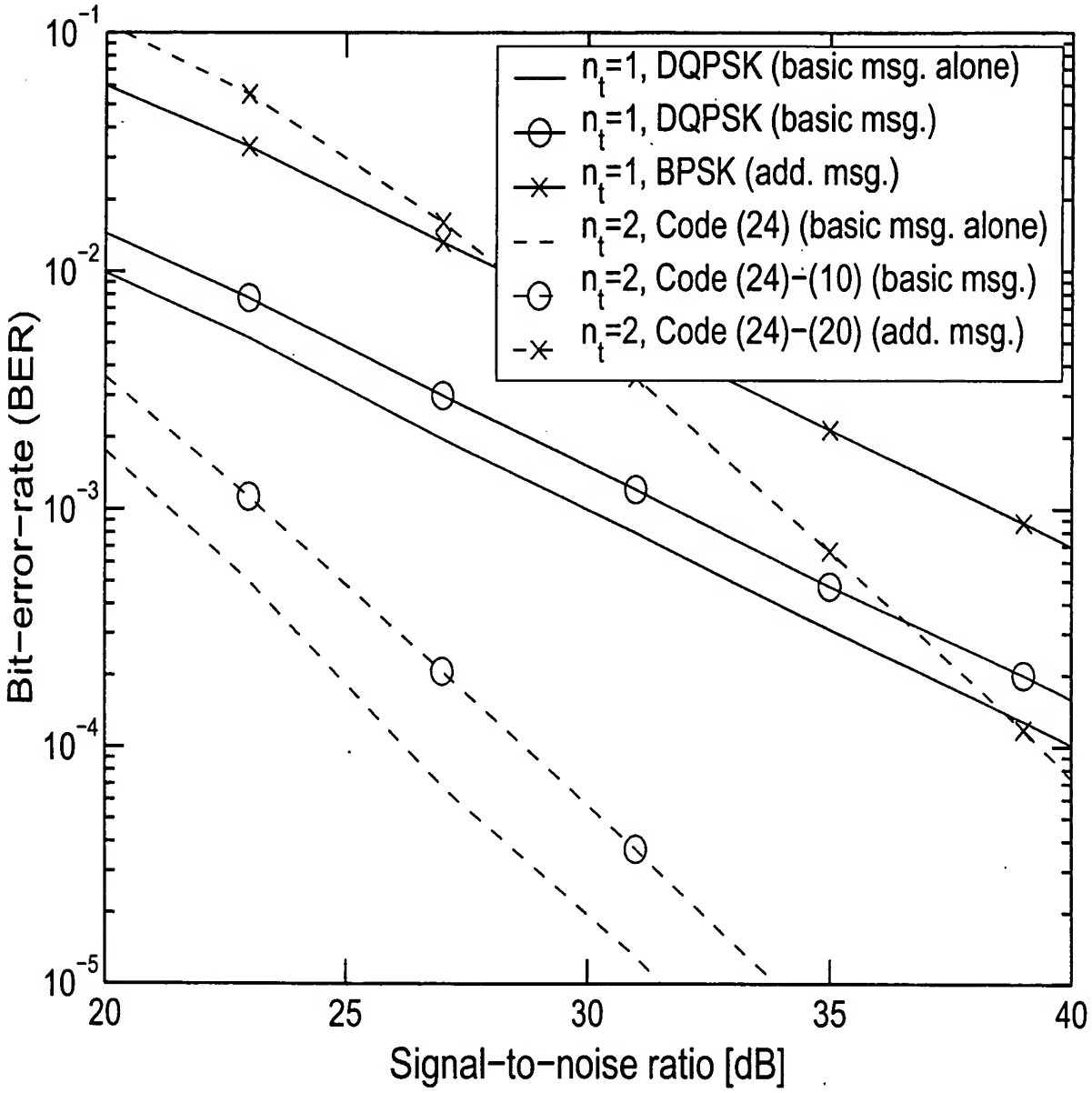

Fig. 5. Empirical BER for the $R_b = 2$, $R_a = 1$ nonuniform space-time constellation. The parameters were $\lambda = 0.078$, $\gamma = 0.018$ (found via optimization of (17)).

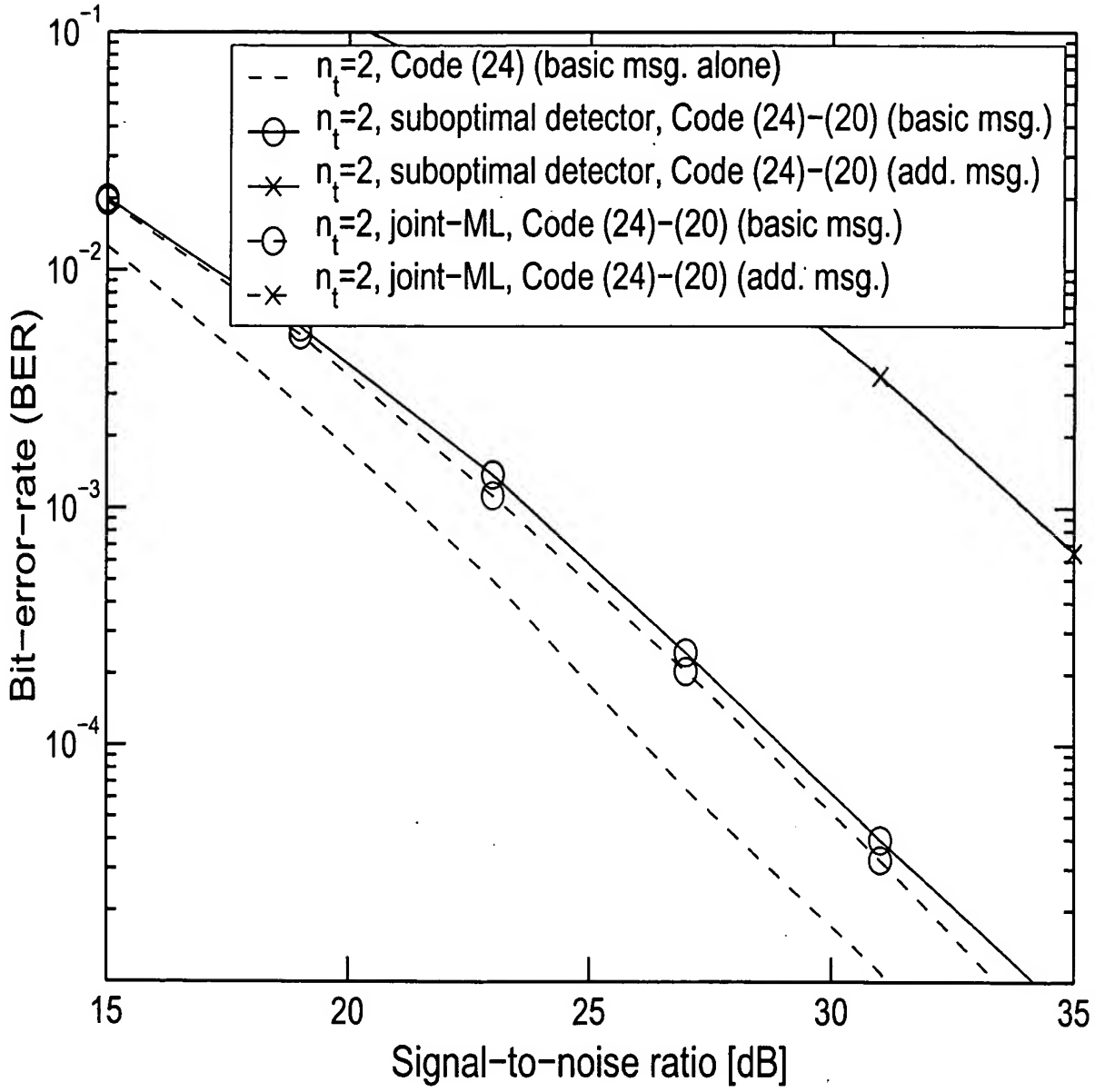

Fig. 6. Monte-Carlo Simulation of the BER performance for the suboptimal receiver (26) and (27) in Section VII, for the joint ML receiver and for the basic message only system with $R_b = 2$, $R_a = 1$ code. The solid and dash curves with "×" overlap.

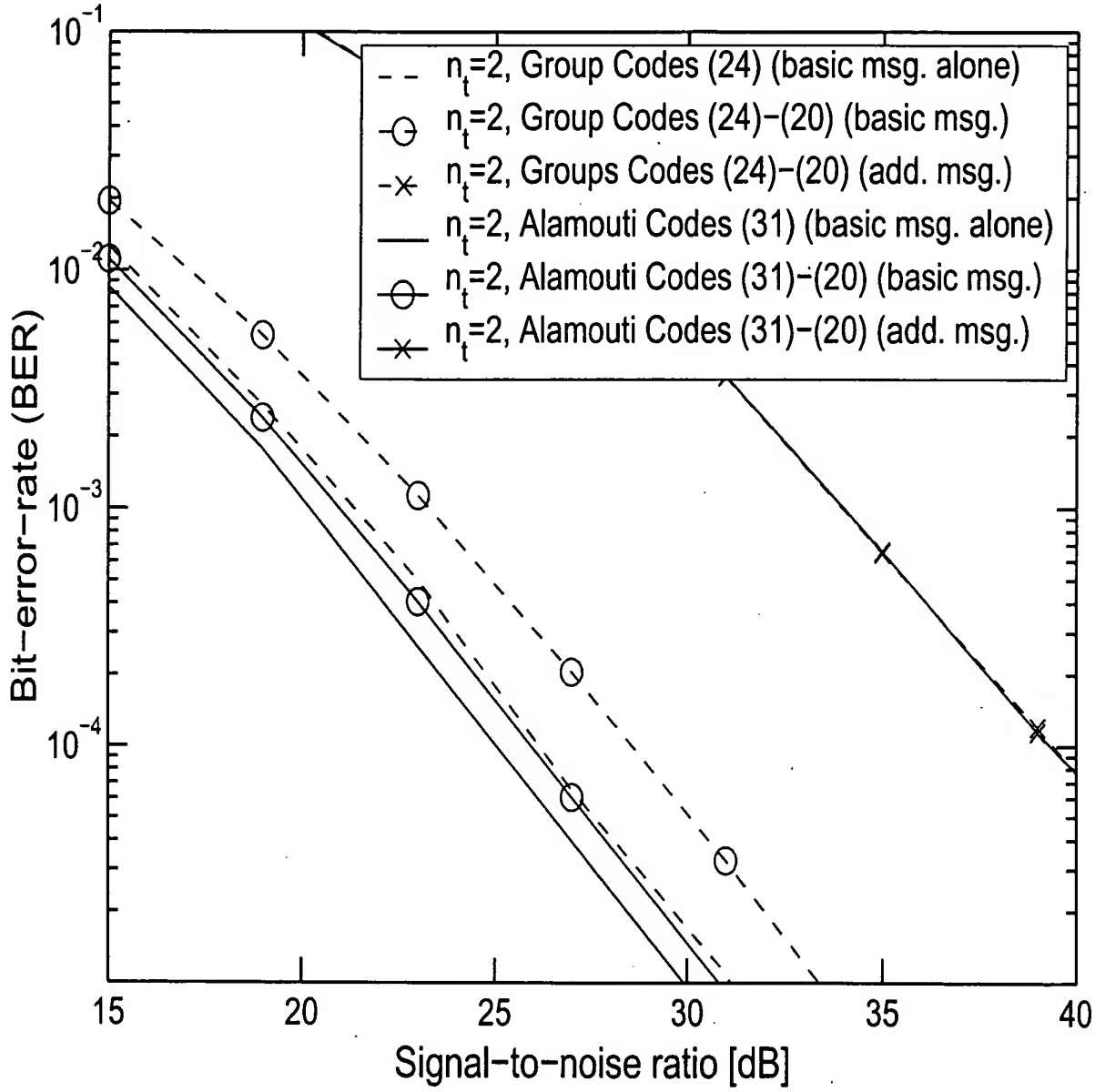

Fig. 7. Monte Carlo Simulation result for the BER performance of $R_b = 2$, $R_a = 1$ system, with differential group codes and differential Alamouti codes. Note that the solid and dash curves with "×" marks overlap.

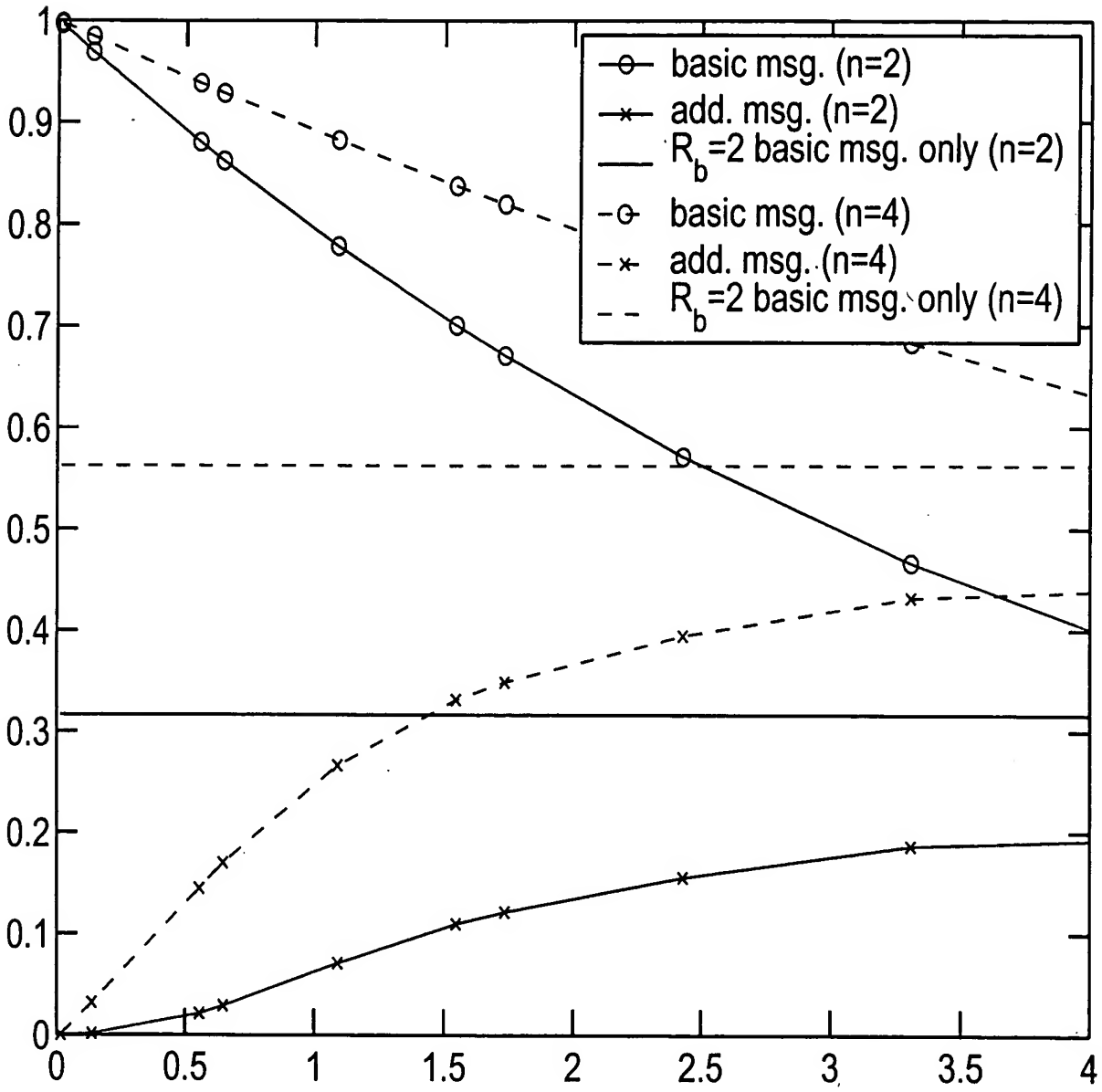

Fig. 8. Outage area for the basic/additional messages as a function of the "tolerable" performance loss, for the $R_b = 1$, $R_a = 1$ code. The horizontal curves represent the relative outage area for the $R_b = 2$ code with a basic message only.

NONUNIFORM SPACE-TIME CODES FOR LAYERED SOURCE CODING

By

WING HIN WONG

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2003

To my family.

ACKNOWLEDGMENTS

First I have to thank my advisor, Dr. Erik G. Larsson. Without his continuous guidance, this thesis could never have been a reality. I am so grateful that I was given the opportunity to work on such a challenging topic.

Of course Dr. John M. Shea and Dr. Tan F. Wong cannot be omitted here. I wish to express my sincere thanks to them for their supporting roles in my committee.

The continuous support from my family is the key component to the success of my graduate study. The gratefulness in my mind cannot be expressed by simple languages.

Last but not least, I wish to say thank you to all my lovely friends in Gainesville, Florida. Their encouragement has helped make this thesis better.

TABLE OF CONTENTS

LIST OF FIGURES

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

NONUNIFORM SPACE-TIME CODES FOR LAYERED SOURCE CODING

By

Wing Hin Wong

May 2003

Chair: Erik G. Larsson
Major Department: Electrical and Computer Engineering

We propose new space-time codes tailored to point-to-multipoint, or broadcast, communications using layered source coding. Our codes can be encoded (and decoded) differentially, and they are based entirely on phase-shift keying. We develop analytical design criteria for the codes, and we discuss the design of optimal and suboptimal receiver structures. We also discuss the relation between our new codes and a differentially encoded Alamouti code. Numerical examples illustrate the performance of our new codes. Convolutional codes are also introduced as a pre-coding to the space-time coding. With some reduction of the data rate, we can significantly increase the error performance in an adverse channel environment. Soft-decision decoding is being used to further improve the performance of the convolutional codes. For completeness, we also include some fundamental concepts of the space-time coding system and the cellular system for the ease of reading for newcomers in this field.

CHAPTER 1
INTRODUCTION

During the last decade, wireless systems with multiple transmit and receive antennas have been studied extensively, and the performance of such systems has been proven to be extremely promising. The propagation channel associated with a system with multiple transmit and receive antennas is sometimes called a *multiple-input multiple-output* (MIMO) channel, and the associated coding and signal processing is referred to as *space-time coding*. Using proper space-time coding, it is possible to use the degrees of freedom of the MIMO channel both to increase the throughput and to counteract fading. For this reason, MIMO technology is believed to become a major cornerstone in many future wireless communication systems.

Loosely speaking, communication links can be classified into two categories: point-to-point links and point-to-multipoint links. In the former case, there is exactly one transmitter and one receiver which communicate with each other at a given time, whereas for a point-to-multipoint link, a transmitted message is aimed at multiple different recipients simultaneously. A cellular communication system with mobile users is an example of a point-to-point communication

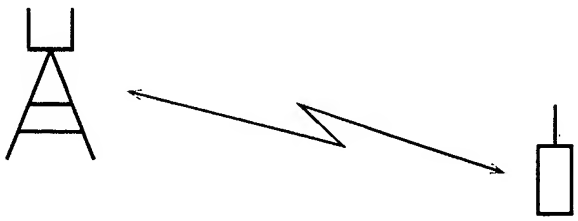

Figure 1–1: A point-to-point link.

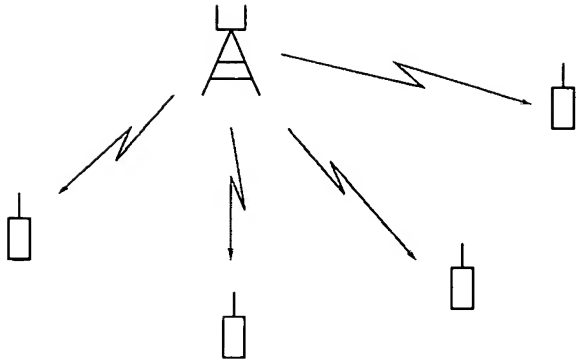

Figure 1–2: A point-to-multipoint link.

system, whereas a radio/TV broadcast is an example of a point-to-multipoint link (therefore point-to-multipoint links are sometimes referred to as broadcast channels).

Point-to-multipoint links are becoming increasingly important. For instance, the introduction of Digital Audio Broadcast (DAB) and High-Definition Television (HDTV) has pioneered a whole new field of digital broadcasting applications. As a further example, it is widely believed that much of the next generation's wireless networking will be based on so-called ad-hoc networks, where it may be necessary for multiple units to listen to one message at the same time. Also, in conventional cell communication system using directional or adaptive antennas, it is sometimes necessary to broadcast a message to the entire cell.

There are two major differences between point-to-point and point-to-multipoint communication links. First, a point-to-point connection can be optimized for a given transmitter-receiver pair. For instance, a cellular system usually employs power control techniques that adjust the transmitted power to minimize the power consumption, reduce the amount of co-channel interference, and at the same time ensure that the received signal strength exceeds a

certain threshold. Second, in contrast to a point-to-point communication link, for a broadcast transmission *all receivers have different capabilities to decode the transmitted message.* This is so because the different receivers experience in general very different radio link qualities (for instance, due to varying propagation conditions). Moreover, since quality can usually be traded for cost, the receivers themselves may have different *inherent* abilities to decode the transmitted message.

The fact that distinct receivers have different capabilities of decoding a message suggests that the transmitted signal should consist of several components which are of different importance for the reconstruction of the message (and therefore have an inherently different vulnerability to transmission errors). This idea has lead to the concept of layered source coding (e.g., [1, 2, 3, 4]), which is now a mature technique employed in many multimedia standards. For instance, the image coding standard JPEG-2000 and the video coding standard MPEG-4 use what is sometimes referred to as "fine granularity scalability," which enables a gradual tradeoff between the error-free data throughput and the quality of the reconstructed image or video sequence [5]. Such progressive source coding methods are already in use in many Internet applications where data rate can be traded for quality, and they are expected to play an instrumental role for the next generation of wireless standards to provide ubiquitous access both to the Internet, and to diverse sources of streaming video and audio.

The topic of this thesis is space-time coding for broadcast channels when layered source coding is used. Previous works on the topic include Memarzadeh et al. [6], which discussed "beamforming" techniques (hence requiring feedback of channel state information) for a multiuser system. In Memarzadeh et al. [6], and two kinds of beamforming techniques are discussed. Zero-forcing beamforming is

introduced which orthogonal codewords are used for each user so that inter-user interference can be fully eliminated just as in an ideal CDMA system, provided that exact channel state information is available at the transmitter. Another way is the single user optimal beamformer, for which no attempt is made to avoid inter-user interference. This is suitable for the scenario when the channel state information at the transmitter is not very accurate, as interference suppression is not effective.

Kuo et al. [7] proposed a layered space-time coding scheme assuming different receivers may have different numbers of receive antennas, and that a receiver with more receive antennas can decode more layers of messages.

In this thesis, we suggest a new layered space-time coding scheme that does not require any channel knowledge at the transmitter, and which is constructed starting from a criterion that attempts to minimize the error rate. Our new codes also satisfy two other important design goals. First, they can be encoded (and detected) differentially. Second, they are entirely based on phase-shift keying (PSK) modulation and consequently the transmitted signal has constant envelope at all times. The codes proposed in this paper can therefore be seen as a multidimensional extension of a transmission technique for single-antenna systems in Pursley et al. [8].

CHAPTER 2
BROADCAST CHANNELS

Broadcast channels are defined as the simultaneous communications between a single source and several receivers.

Early works include Cover et al. [9] and Bergmans et al. [10]. In Cover et al. [9], They discuss the nature of broadcast channel from the perspective of information theory, and derives upper bounds of the capacity of broadcast channels of different types. It is proved that the maximum data rate of each channels are jointly achievable by proper channel coding. By superimposing side information to the common message intended for all receivers, different data rates are possible for different channels. This idea is similar to the nonuniform codes which we will discuss later on.

Due to the different capabilities of each receivers, and indeed the capabilities of each receivers are varying from time to time (i.e., due to channel fading or fluctuating noise level), the theoretical maximum capacity of a broadcast channel is not easy to achieve in a practical system.

We have no intention to compare the performance of our nonuniform codes with the theoretical channel capacity in this thesis, but as our nonuniform codes contain data in multiple layers, we believe that we are in the right direction that our nonuniform codes can increase the total throughput of a broadcast system.

CHAPTER 3
LAYERED SOURCE CODING

In this chapter we briefly introduce the concept of layered source coding. JPEG-2000 is taken as the example here.

JPEG is an image compression standard developed by the Joint Photographic Experts Group [11]. For the encoding, the pixel bit map information of a whole image is divided into 8×8 pixel blocks. Each 8×8 pixel block then undergoes a discrete cosine transformation (DCT). This helps separate the image into parts (or spectral sub-bands) of different importance (with respect to the image's visual quality). The DCT is similar to the discrete Fourier transform as it transforms a signal or image from the spatial domain to the frequency domain. For most images, much of the signal energy lies at low frequencies, whereas the higher frequency components are often small and usually they can be neglected with little visible distortion. Thus the DCT indeed separates the original image data into separate layers of different importance: the low frequency components are important for the reconstruction of the image; the higher frequency terms are not that important although they enhance the visual quality of the image. From this observation, we infer that the low frequency parts of data must be sent through a reliable data channel with very good bit-error-rate (BER) performance whereas the high frequency parts can be sent through a less reliable channel as the occurrence of bit errors will only deteriorate the quality to a negligible extent.

With layered source coding (for instance as mentioned above, JPEG-2000 and MPEG-4), more capable receivers, for example with higher signal-to-noise-ratio (SNR), can achieve a higher data rate by decoding the messages contained in all layers while the less capable receivers can only decode the message in the bottom layer. Thus the "most important" information should be conveyed via the bottom layer while "less important" information can be transmitted via higher layer(s).

An example of a modulation scheme designed for layered source coding is nonuniform phase-shift-keying modulation (PSK), which was first introduced by Pursley and Shea [8] (see also Pursley et al. [12, 13]). The idea is intuitively appealing. Starting with a standard uniform PSK constellation, which contains data for the "*basic*" layer, a nonuniform PSK code constellation is obtained by adding a small additional phase shift α to each original constellation point, which contains the information for the "*additional*" layer. This is illustrated in Figure 3–1, when the *basic* message is conveyed via a QPSK constellation. A capable receiver with high SNR can distinguish among all eight constellation points which means that both the *basic* and the *additional* layer messages can be decoded, whereas to a less capable receiver the constellation may appear like a blurred QPSK constellation and hence it may only be able to distinguish between the different "large" phase shifts; thus only the *basic* message can be detected accurately. By such a construction the error rates associated with the basic and the additional message are different. Consequently, the additional message can carry information that is of less importance for the reconstruction of the transmitted message than the basic message is. The error probability of the additional message can be easily adjusted by choosing different values of the additional phase shift α. Obviously, a larger α results in a better performance

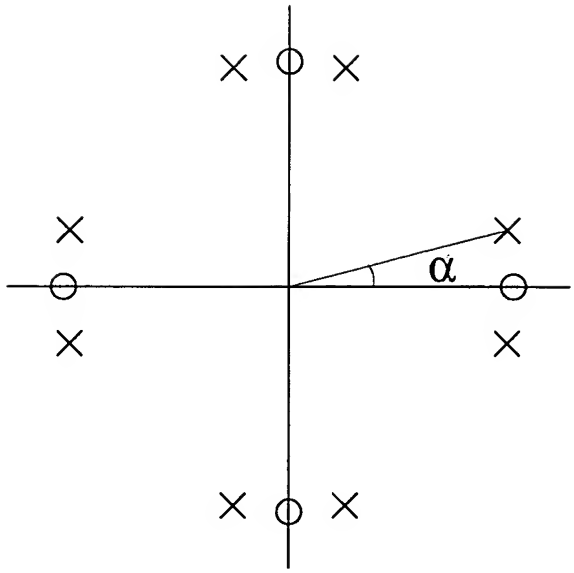

Figure 3–1: A nonuniform 8-PSK constellation obtained from a standard "uniform" QPSK constellation by splitting each original constellation point "o" into a pair of new points "×".

for the additional message at the cost of a deterioration of the performance for the basic message. The main virtue of this nonuniform PSK encoding scheme is that the signaling envelope is constant while multi-layered messages can be transmitted with different and adjustable error performance.

CHAPTER 4
SPACE-TIME CODING

4.1 Error Performance of STBC Systems

A conventional wireless system with a single transmit and receive antenna suffers from a large performance degradation due to the fading of the radio channel. Fading is the fluctuation in received signal level. It is mainly due to changes in the propagation environment, such as weather and moving reflective objects/obstructions.

Fading can be classified as *fast fading* or *slow fading*, depending on how rapidly the channel impulse response changes. For a fast fading channel, the channel impulse response changes rapidly within one symbol duration, which means that the coherence time of the channel T_c (which is a measure of the time duration that the channel impulse response is quasi-static) is smaller than the symbol period of the transmitted signal

$$T_s > T_c \tag{4.1}$$

For a slow fading channel, the channel impulse response changes very slowly compared with one symbol duration, which means that the channel can be seen as quasi-static for many symbol durations, that is

$$T_s \ll T_c \tag{4.2}$$

Figure 4–1 shows an example of the fast fading of the gain of a single radio channel. The two curves represent two channels encounter independent fast fading. We can see that there is a high risk that a single channel may encounter

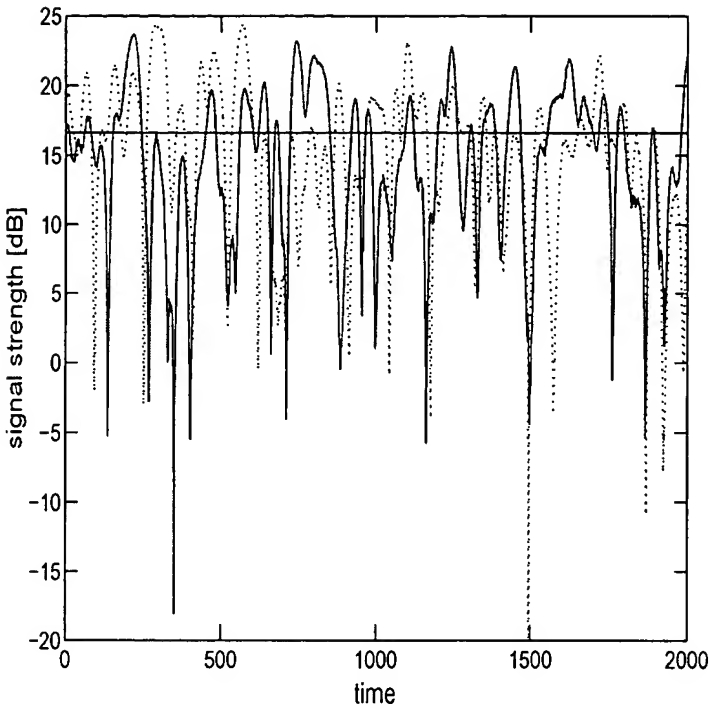

Figure 4–1: Received signal level may fluctuate vastly for a single radio channel.

a deep fade at a certain time. During deep fade the transmission quality may be very poor or the transmission may simply be impossible.

To counteract the fading radio environment a multiple transmit and receive antenna system (or MIMO) is one attractive possibility. The advantage of a multiple transmit and receive antenna is not difficult to comprehend intuitively: There are more than one channel which are not fully dependent, so the possibility that all of them encounter a deep fading at the same time is small; thus the deep fading problem may be solved. Figure 4–2 shows the total channel gain (denote the individual channel gains be h_1 and h_2, total channel gain is then equal to $|h_1|^2 + |h_2|^2$) of two independent fading channels. It can be seen that the risk of the total channel gain encountering a deep fade is largely reduced.

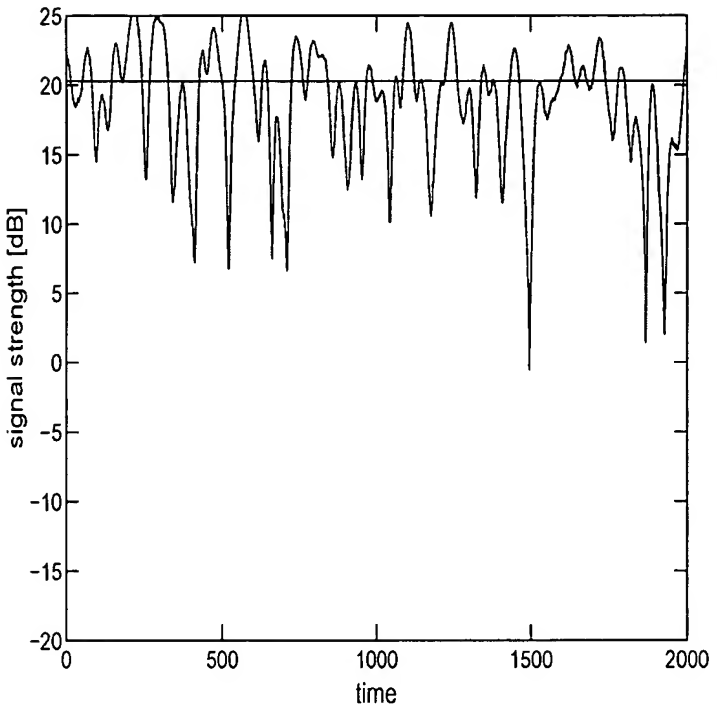

Figure 4–2: Total channel gain of two independent fading channels.

The corresponding encoding scheme for such a multi transmit and receive antenna system is called space-time coding. This is an emerging topic that now attracts many researchers' attention since its introduction about a decade ago.

The main virtue of such a multi transmit and receive antenna system is that higher diversity order can be achieved. The diversity order can be defined implicitly by

$$\text{Bit-Error-Rate} \propto (\text{SNR})^{-\text{diversity}} \tag{4.3}$$

which means that the bit-error-rate decreases as fast as the Signal-to-Noise-Ratio (SNR) to the power of minus the diversity order. For a high SNR the advantage of a high diversity order can be substantial.

Consider a MIMO system with n_t transmit antennas, n_r receive antennas (see Figure 4–3), and assume for simplicity that the propagation channel is

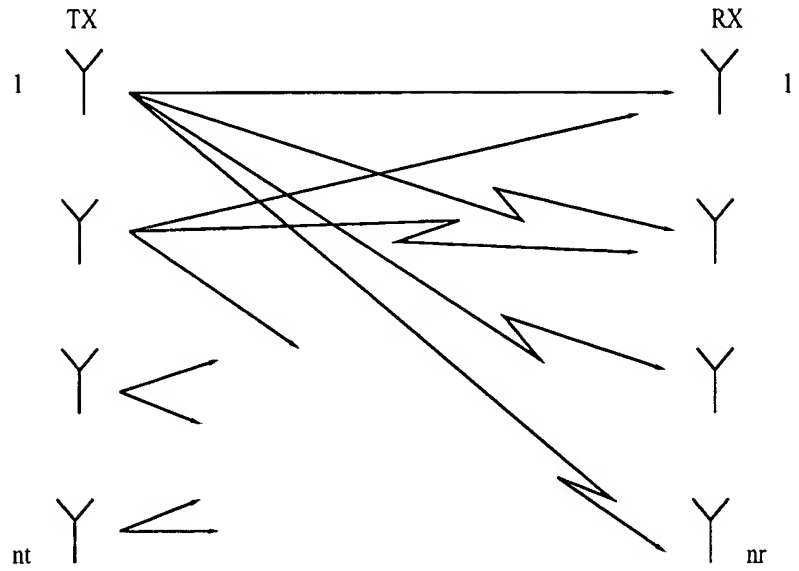

Figure 4–3: A wireless link with multi transmit and receive antennas.

linear, time-invariant and frequency flat. Let $\boldsymbol{H}$ be an $n_r \times n_t$ matrix whose (m,n)th element contains the channel gain between transmit antenna n and receive antenna m, and suppose that a code matrix $\boldsymbol{X}$ of dimension $n_t \times N$ is taken from a matrix constellation $\mathcal{X}$ and transmitted by sending its N columns via the n_t antennas during N time epochs. If the signal received during the same N time intervals is arranged in an $n_r \times N$ matrix $\boldsymbol{Y}$, then

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E} \tag{4.4}$$

where $\boldsymbol{E}$ is an $n_r \times N$ matrix of noise. Throughout this thesis, we shall make the somewhat standard assumption that the elements of $\boldsymbol{H}$ and $\boldsymbol{E}$ are independent and zero-mean complex Gaussian with variances ρ^2 and σ^2, respectively; hence the channel is Rayleigh fading.

The Shannon channel capacity of such a MIMO system is a well-known result, and it can be shown that (for details see [14] chapter 3 and the references

therein)

$$C(\boldsymbol{H}) = B \cdot \log_2 |\boldsymbol{I} + \frac{1}{\sigma^2}\boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^H| \tag{4.5}$$

where B is the bandwidth of the channel in Hz and $\boldsymbol{P}$ is the covariance matrix of the sequence of transmitted data vectors

$$\boldsymbol{P} = E[\boldsymbol{x}_n \boldsymbol{x}_n^H] \tag{4.6}$$

If the transmitted signals are uncorrelated at different transmit antennas and we use unit power at each transmit interval, then

$$\boldsymbol{P} = \frac{1}{n_t} \cdot \boldsymbol{I} \tag{4.7}$$

(4.5) reduced to

$$C(\boldsymbol{H}) = B \cdot \log_2 |\boldsymbol{I} + \frac{1}{n_t\sigma^2}\boldsymbol{H}\boldsymbol{H}^H| \tag{4.8}$$

Clearly, the channel capacity depends on the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^H$.

Given $\boldsymbol{Y}$, detection of $\boldsymbol{X}$ in a maximum-likelihood (ML) sense amounts to minimizing the Euclidian distance

$$\|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\| \tag{4.9}$$

with respect to $\boldsymbol{X} \in \mathcal{X}$, as well as $\boldsymbol{H}$ (unless it is known). The average (over $\boldsymbol{H}$) probability that the ML detector mistakes a transmitted code matrix $\boldsymbol{X}_0$ for an incorrect (and different) code matrix $\boldsymbol{X} \neq \boldsymbol{X}_0$ is called the pairwise error probability For completeness a brief derivation of this probability is provided here for both coherent and noncoherent detection.

Coherent: Suppose that $\boldsymbol{X}_0$ is the true transmitted code matrix, thus $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X}_0 + \boldsymbol{E}$. Then the probability that an incorrect decision in favor of a

different code matrix $\boldsymbol{X} \neq \boldsymbol{X}_0$ occurs is

$$\begin{aligned} P(\text{error}|\boldsymbol{H}) =& P\big(\|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}\|^2 < \|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}_0\|^2\big|\boldsymbol{H}\big) \\ =& P\big(\|\boldsymbol{H}\tilde{\boldsymbol{X}}+\boldsymbol{E}\|^2 < \|\boldsymbol{E}\|^2\big|\boldsymbol{H}\big) \\ =& P\Big(2\,\mathrm{Re}\,\mathrm{Tr}\{\boldsymbol{E}^H\boldsymbol{H}\tilde{\boldsymbol{X}}\} > \|\boldsymbol{H}\tilde{\boldsymbol{X}}\|^2\big|\boldsymbol{H}\Big) \\ \leq& \exp\Big(-\frac{\|\boldsymbol{H}\tilde{\boldsymbol{X}}\|^2}{4\sigma^2}\Big) \end{aligned} \tag{4.10}$$

where $\tilde{\boldsymbol{X}} = \boldsymbol{X}_0 - \boldsymbol{X}$, and $|\cdot|$ denotes the determinant of a matrix. In (4.10), the second equality follows by some matrix algebra and the last step is a consequence of the fact that $2\,\mathrm{Re}\,\mathrm{Tr}\{\boldsymbol{E}^*\boldsymbol{H}\tilde{\boldsymbol{X}}\}$ is zero-mean Gaussian with variance $2\sigma^2\|\boldsymbol{H}\tilde{\boldsymbol{X}}\|^2$ together with the Chernoff bound. Under the assumptions made, the probability density function (p.d.f.) for $\boldsymbol{H}$ is

$$p(\boldsymbol{H}) = \frac{1}{\pi^{n_r n_t}\rho^{2n_r n_t}}\exp\Big(-\frac{1}{\rho^2}\|\boldsymbol{H}\|^2\Big) \tag{4.11}$$

Averaging (4.10) over $\boldsymbol{H}$ using (4.11), we get after some calculations

$$\begin{aligned} P(\text{error}) =& \int d\boldsymbol{H}\; P(\text{error}|\boldsymbol{H})p(\boldsymbol{H}) \\ \leq& (\pi\rho^2)^{-n_r n_t}\int d\boldsymbol{H}\;\exp\Big(-\frac{1}{4\sigma^2}\|\boldsymbol{H}\tilde{\boldsymbol{X}}\|^2 - \frac{1}{\rho^2}\|\boldsymbol{H}\|^2\Big) \\ =& (\pi\rho^2)^{-n_r n_t}\int d\boldsymbol{H}\;\exp\Big(-\mathrm{Tr}\Big\{\boldsymbol{H}\cdot\frac{1}{\rho^2}\Big(\frac{\rho^2}{4\sigma^2}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^H + \boldsymbol{I}\Big)\cdot\boldsymbol{H}^H\Big\}\Big) \\ =& \Big|\frac{\rho^2}{4\sigma^2}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^H + \boldsymbol{I}\Big|^{-n_r} \leq |(\boldsymbol{X}-\boldsymbol{X}_0)(\boldsymbol{X}-\boldsymbol{X}_0)^H|^{-n_r}\Big(\frac{\rho^2}{4\sigma^2}\Big)^{-n_r n_t} \end{aligned} \tag{4.12}$$

Noncoherent: If $\boldsymbol{H}$ is unknown, we can obtain a detection rule that depends only on $\boldsymbol{X}$. Consider the problem of minimizing (4.9)

$$\min_{\boldsymbol{X}\in\mathcal{X},\boldsymbol{H}} \quad \|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}\|^2 \tag{4.13}$$

Treating $\boldsymbol{H}$ as a deterministic unknown, we can minimize this likelihood function with respect to $\boldsymbol{H}$. Notice that

$$\begin{aligned}
\|\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}\|^2 &= \left\|\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H\right\|^2 \\
&= \left\|(\Pi_{\boldsymbol{X}^H}+\Pi^{\perp}_{\boldsymbol{X}^H})(\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H)\right\|^2 \\
&= \left\|\Pi_{\boldsymbol{X}^H}(\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H)\right\|^2+\left\|\Pi^{\perp}_{\boldsymbol{X}^H}(\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H)\right\|^2 \\
&\quad \underbrace{+2\,\mathrm{Re}\left(\mathrm{Tr}\left\{(\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X})\Pi_{\boldsymbol{X}^H}\cdot\Pi^{\perp}_{\boldsymbol{X}^H}(\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H)\right\}\right)}_{0} \\
&= \left\|\Pi_{\boldsymbol{X}^H}\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H\right\|^2+\left\|\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{Y}^H\right\|^2
\end{aligned} \tag{4.14}$$

Here $\Pi_{\boldsymbol{X}^H}=\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1}\boldsymbol{X}$ is the orthogonal projector of $\boldsymbol{X}^H$. The function in (4.14) is minimized when

$$\begin{aligned}
&\Pi_{\boldsymbol{X}^H}\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H=0 \\
\Longleftrightarrow\;&\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1}\boldsymbol{X}\boldsymbol{Y}^H-\boldsymbol{X}^H\boldsymbol{H}^H=0 \\
\Longleftrightarrow\;&\boldsymbol{X}^H\left((\boldsymbol{X}\boldsymbol{X}^H)^{-1}\boldsymbol{X}\boldsymbol{Y}^H-\boldsymbol{H}^H\right)=0
\end{aligned} \tag{4.15}$$

If $\boldsymbol{X}$ has full rank then the above leads to the result (see, e.g., [15] and [14] appendix B)

$$\boldsymbol{H}=\boldsymbol{Y}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1} \tag{4.16}$$

(It may be argued that if we know some sort of statistical properties of $\boldsymbol{H}$, we should treat $\boldsymbol{H}$ as a stochastic variable when it enters the likelihood function. This is intuitive since we should incorporate all known information into the detector to get the best result. But in Larsson et al. [16] (see also Larsson et al. [14] exercise 9.7) it is shown that both approaches indeed give the same result. In this paper we will stick to the deterministic channel approach for simplicity.)

Inserting (4.16) into (4.13) yields

$$\begin{aligned} &\min_{\boldsymbol{X}\in\mathcal{X}} \ \|\boldsymbol{Y}-\boldsymbol{Y}\boldsymbol{X}^H(\boldsymbol{X}\boldsymbol{X}^H)^{-1}\boldsymbol{X}\|^2 \\ \Longleftrightarrow &\min_{\boldsymbol{X}\in\mathcal{X}} \ \|\boldsymbol{Y}\Pi^{\perp}_{\boldsymbol{X}^H}\|^2 \\ \Longleftrightarrow &\max_{\boldsymbol{X}\in\mathcal{X}} \ \|\Pi_{\boldsymbol{X}^H}\boldsymbol{Y}^H\|^2 \end{aligned} \tag{4.17}$$

Let

$$\Delta_{\Pi} = \Pi_{\boldsymbol{X}_0^H} - \Pi_{\boldsymbol{X}^H} \tag{4.18}$$

Note that the matrix Δ_{Π} is Hermitian, but *not* necessarily positive definite. Also note that

$$\begin{aligned} \Delta_{\Pi}\boldsymbol{X}_0^H &= (\Pi_{\boldsymbol{X}_0^H} - \Pi_{\boldsymbol{X}^H})\boldsymbol{X}_0^H \\ &= \boldsymbol{X}_0^H - \Pi_{\boldsymbol{X}^H}\boldsymbol{X}_0^H \\ &= \Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H \end{aligned} \tag{4.19}$$

Hence, since $\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H$ has full rank by assumption, the matrix $\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H$ is positive definite.

The probability that the ML detector makes a mistake is

$$\begin{aligned} &P(\text{error}|\boldsymbol{H}) = P\Big(\|\Pi_{\boldsymbol{X}^H}\boldsymbol{Y}^H\|^2 > \|\Pi_{\boldsymbol{X}_0^H}\boldsymbol{Y}^H\|^2\Big|\boldsymbol{H}\Big) \\ =&P\Big(\operatorname{Tr}\left\{(\boldsymbol{H}\boldsymbol{X}_0+\boldsymbol{E})\Delta_{\Pi}(\boldsymbol{H}\boldsymbol{X}_0+\boldsymbol{E})^H\right\} < 0\Big|\boldsymbol{H}\Big) \\ =&P\Big(2\operatorname{Re}\operatorname{Tr}\left\{\boldsymbol{H}\boldsymbol{X}_0\Delta_{\Pi}\boldsymbol{E}^H\right\} \\ &+\operatorname{Tr}\{\boldsymbol{E}\Delta_{\Pi}\boldsymbol{E}^H\} < -\operatorname{Tr}\{\boldsymbol{H}\boldsymbol{X}_0\Delta_{\Pi}\boldsymbol{X}_0^H\boldsymbol{H}^H\}\Big|\boldsymbol{H}\Big) \\ =&P\Big(2\operatorname{Re}\operatorname{Tr}\left\{\boldsymbol{H}\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{E}^H\right\} \\ &+\operatorname{Tr}\{\boldsymbol{E}\Delta_{\Pi}\boldsymbol{E}^H\} < -\|\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H\boldsymbol{H}^H\|^2\Big|\boldsymbol{H}\Big) \end{aligned} \tag{4.20}$$

where in the last step we used the fact that $\Pi^{\perp}_{\boldsymbol{X}^H}\Pi^{\perp}_{\boldsymbol{X}^H} = \Pi^{\perp}_{\boldsymbol{X}^H}$.

The term $\boldsymbol{E}\boldsymbol{\Delta}_{\Pi}\boldsymbol{E}^H$ is of higher order and may be neglected in an asymptotic analysis. In particular, for a given $\boldsymbol{H}$, the variance of its elements is of the order σ^4, whereas the elements of $\boldsymbol{H}\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{E}^H$ have a variance of the order σ^2. Since

$$\frac{dQ(x)}{dx} = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \leq \frac{1}{\sqrt{2\pi}} \leq \infty \tag{4.21}$$

Thus a small finite error in variance can only caused a small finite error in the Q-function also, so the additional small variance caused by this higher order term can be neglected.

By the Chernoff bound we have

$$\begin{aligned} &P\Big(2\,\mathrm{Re}\,\mathrm{Tr}\{\boldsymbol{H}\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{E}^H\} < -\|\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H\boldsymbol{H}^H\|^2 \big| \boldsymbol{H}\Big) \\ \lesssim &\exp\Big(-\frac{\|\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H\boldsymbol{H}^H\|^2}{4\sigma^2}\Big) \end{aligned} \tag{4.22}$$

Hence, by averaging in a way similar to (4.23), we get

$$\begin{aligned} E_{\boldsymbol{H}}[P(\text{error})] &= \int d\boldsymbol{H}\ P(\text{error}|\boldsymbol{H})p(\boldsymbol{H}) \\ &\lesssim (\pi\rho^2)^{-n_r n_t}\int d\boldsymbol{H}\ \exp\Big(-\frac{1}{4\sigma^2}\|\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H\boldsymbol{H}^H\|^2 - \frac{1}{\rho^2}\|\boldsymbol{H}\|^2\Big) \\ &= (\pi\rho^2)^{-n_r n_t}\int d\boldsymbol{H}\ \exp\Big(-\mathrm{Tr}\Big\{\boldsymbol{H}\cdot\frac{1}{\rho^2}(\frac{\rho^2}{4\sigma^2}\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H + \boldsymbol{I})\cdot\boldsymbol{H}^H\Big\}\Big) \\ &= \Big|\frac{\rho^2}{4\sigma^2}\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H + \boldsymbol{I}\Big|^{-n_r} \leq |\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H|^{-n_r}\cdot\Big(\frac{\rho^2}{4\sigma^2}\Big)^{-n_r n_t} \end{aligned} \tag{4.23}$$

To summarize, the error bounds for coherent and noncoherent detection in a STBC system are the following

$$E\left[P(\boldsymbol{X}_0 \to \boldsymbol{X})\right]\Big|_{\text{coherent}} \leq \left|(\boldsymbol{X}-\boldsymbol{X}_0)(\boldsymbol{X}-\boldsymbol{X}_0)^H\right|^{-n_r}\cdot\Big(\frac{\rho^2}{4\sigma^2}\Big)^{-n_r n_t} \tag{4.24}$$

and

$$E\left[P(\boldsymbol{X}_0 \to \boldsymbol{X})\right]\Big|_{\text{noncoherent}} \leq \left|\boldsymbol{X}_0\Pi^{\perp}_{\boldsymbol{X}^H}\boldsymbol{X}_0^H\right|^{-n_r}\cdot\Big(\frac{\rho^2}{4\sigma^2}\Big)^{-n_r n_t} \tag{4.25}$$

A special subclass of space-time codes is the linear *orthogonal space-time block codes (OSTBC)*. The case when OSTBC applied to a 2-TX system turns out to be the well-known Alamouti code [17].

The main virtue of OSTBC is that they achieve a diversity of order $n_r n_t$ for a system with n_t transmit and n_r receive antennas, at an almost negligible computational cost. Specifically, for such a system, a set of n_s symbols $\{s_1, \ldots, s_{n_s}\}$ are encoded into an $n_t \times N$ matrix that has the following structure

$$\boldsymbol{X}_{\text{OSTBC}} = \sum_{n=1}^{n_s} \left(s_n \boldsymbol{A}_n + s_n^* \boldsymbol{B}_n \right) \tag{4.26}$$

The columns of X_{OSTBC} are transmitted via the n_t transmit antennas during N time intervals. In (4.26), $\{\boldsymbol{A}_n, \boldsymbol{B}_n\}$ are matrices chosen such that

$$\boldsymbol{X}_{\text{OSTBC}} \boldsymbol{X}_{\text{OSTBC}}^H = \sum_{n=1}^{n_s} |s_n|^2 \cdot \boldsymbol{I} \tag{4.27}$$

for all $\{s_n\}$. In general, the role of the orthogonality condition (4.27) and the associated design of $\{\boldsymbol{A}_n, \boldsymbol{B}_n\}$ is relatively well understood by now; for instance, (4.27) implies that the symbols $\{s_n\}$ can be detected independently of each other [14], so with the use of OSTBC, the MIMO channel can be decoupled into n_s number of single additive-white-gaussian-noise (AWGN) channels (See Figure 4–4).

4.2 Differential Modulation for MIMO Systems

Space-time modulation matrices that can be encoded *differentially* are sometimes of special interest since such codes can easily be demodulated noncoherently. Differential codes with uniform error probabilities have been studied by many authors (see, e.g., [18, 19, 20, 21, 22] for some prominent examples) and some of them can be seen as an extension of differential PSK to MIMO systems. In general, if t is the time index and if $\{\boldsymbol{U}(t)\}$ is a sequence of (square)

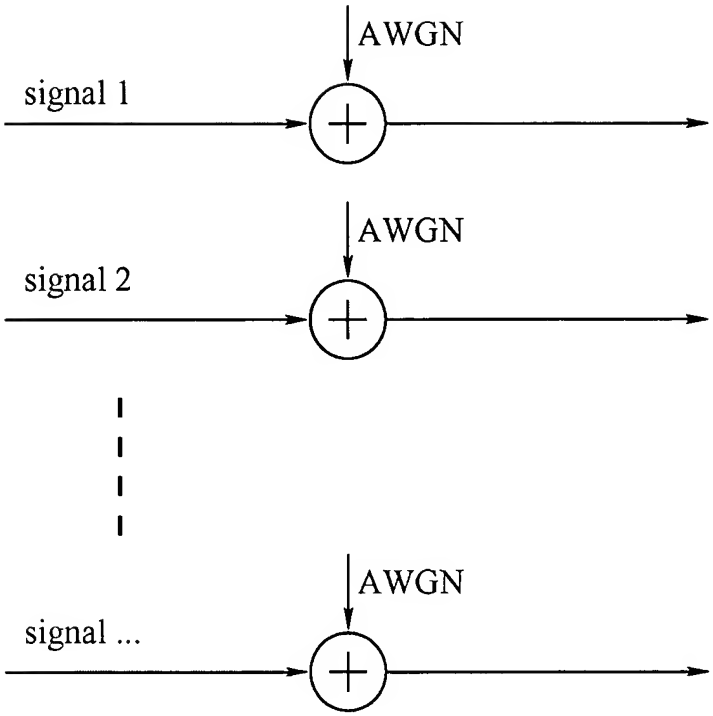

Figure 4–4: OSTBC decouples a MIMO channel into n_s number of AWGN channels.

information-bearing matrices, then differential encoding obtains the transmitted code matrix $\boldsymbol{X}(t)$ at time t via

$$\boldsymbol{X}(t) = \boldsymbol{X}(t-1)\boldsymbol{U}(t) \tag{4.28}$$

where $\boldsymbol{X}(t-1)$ is the matrix transmitted at time $t-1$. Such an encoding usually is only meaningful under certain circumstances, for instance if $\boldsymbol{U}(t)$ is unitary (in which case $\boldsymbol{X}(t)$ becomes unitary for all t as well, given a unitary "initial matrix" $\boldsymbol{X}(0)$). By considering two (or more) received matrices $\boldsymbol{Y}(t)$ and $\boldsymbol{Y}(t-1)$ simultaneously, noncoherent detection is possible. For example, if we concatenate the matrices received at time $t-1$ and t, and assume that the channel $\boldsymbol{H}$ remains constant over these two time intervals (to within practical accuracy), we can write (using (4.4) and (4.28))

$$\begin{aligned}
&\begin{bmatrix}\boldsymbol{Y}(t-1) & \boldsymbol{Y}(t)\end{bmatrix} \\
=&\begin{bmatrix}\boldsymbol{H}\boldsymbol{X}(t-1) & \boldsymbol{H}\boldsymbol{X}(t)\end{bmatrix} + \begin{bmatrix}\boldsymbol{E}(t-1) & \boldsymbol{E}(t)\end{bmatrix} \\
=&\begin{bmatrix}\boldsymbol{H}\boldsymbol{X}(t-1) & \boldsymbol{H}\boldsymbol{X}(t-1)\boldsymbol{U}(t)\end{bmatrix} + \begin{bmatrix}\boldsymbol{E}(t-1) & \boldsymbol{E}(t)\end{bmatrix} \\
=&\boldsymbol{H}\boldsymbol{X}(t-1)\cdot\begin{bmatrix}\boldsymbol{I} & \boldsymbol{U}(t)\end{bmatrix} + \begin{bmatrix}\boldsymbol{E}(t-1) & \boldsymbol{E}(t)\end{bmatrix}
\end{aligned} \tag{4.29}$$

where $\boldsymbol{H}\boldsymbol{X}(t-1)$ can be seen as an unknown "effective" channel matrix and $\begin{bmatrix}\boldsymbol{I} & \boldsymbol{U}(t)\end{bmatrix}$ is an effective code matrix. Provided that $\begin{bmatrix}\boldsymbol{I} & \boldsymbol{U}(t)\end{bmatrix}$ is such that the corresponding determinant in (4.25) is nonzero, noncoherent demodulation is possible. By applying the decision rule (4.17) to the data model (4.29), the

detection rule for differential demodulation becomes

$$\begin{aligned}
&\min_{\boldsymbol{U}(t)\in\mathcal{X}} \left\| \left[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\right] \cdot \Pi^{\perp}_{[\boldsymbol{I}\ \boldsymbol{U}(t)]^H} \right\|^2 \\
\Longleftrightarrow &\min_{\boldsymbol{U}(t)\in\mathcal{X}} \mathrm{Tr}\left\{ \left[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\right] \cdot \Pi^{\perp}_{[\boldsymbol{I}\ \boldsymbol{U}(t)]^H} \cdot \begin{bmatrix} \boldsymbol{Y}^H(t-1) \\ \boldsymbol{Y}^H(t) \end{bmatrix} \right\} \\
\Longleftrightarrow &\min_{\boldsymbol{U}(t)\in\mathcal{X}} \mathrm{Tr}\left\{ \left[\boldsymbol{Y}(t-1) \quad \boldsymbol{Y}(t)\right] \cdot \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{U}(t) \\ -\boldsymbol{U}^H(t) & \boldsymbol{I} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{Y}^H(t-1) \\ \boldsymbol{Y}^H(t) \end{bmatrix} \right\} \\
\Longleftrightarrow &\max_{\boldsymbol{U}(t)\in\mathcal{X}} \mathrm{Re}\left\{ \mathrm{Tr}\left\{\boldsymbol{U}(t)\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1)\right\} \right\}
\end{aligned} \tag{4.30}$$

where $\mathrm{Tr}\{\cdot\}$ stands for the trace of a matrix and where we used the fact that the information-bearing matrix $\boldsymbol{U}(t)$ is unitary. The associated average error probability, i.e., the probability that $\boldsymbol{U}^{(0)}$ is incorrectly detected as $\boldsymbol{U}$ for this differential detection scheme can be bounded by using (4.25)

$$\begin{aligned}
E\left[P(\boldsymbol{U}^{(0)} \to \boldsymbol{U})\right] &\le \left| \left[\boldsymbol{I} \quad \boldsymbol{U}^{(0)}\right] \cdot \Pi^{\perp}_{[\boldsymbol{I}\ \boldsymbol{U}]^H} \cdot \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{U}^{(0)H} \end{bmatrix} \right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \\
&= \left| \boldsymbol{I} - \frac{1}{2}(\boldsymbol{U}\boldsymbol{U}^{(0)H} + \boldsymbol{U}^{(0)}\boldsymbol{U}^H) \right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t}
\end{aligned} \tag{4.31}$$

For simplicity hereafter the time index t of $\boldsymbol{U}(t)$ is omitted whenever no confusion can occur.

4.3 Nonuniform Space-Time Codes

Inspired by the nonuniform PSK codes discussed in Section 3, our new nonuniform space-time codes are based on differential encoding of the product of a unitary code matrix $\boldsymbol{U}_{\mathrm{b}} \in \mathcal{X}_{\mathrm{b}}$ associated with a basic message, and another unitary code matrix $\boldsymbol{U}_{\mathrm{a}} \in \mathcal{X}_{\mathrm{a}}$ corresponding to an additional message. Thus the transmitted code matrix at time t is given by (cf. (4.28))

$$\boldsymbol{X}(t) = \boldsymbol{X}(t-1)\boldsymbol{U}_{\mathrm{b}}(t)\boldsymbol{U}_{\mathrm{a}}(t) \tag{4.32}$$

Using (4.31) along with the fact that $\boldsymbol{U}_{\mathrm{b}}\boldsymbol{U}_{\mathrm{a}}$ is unitary, we find that the average probability that a transmitted message pair $(\boldsymbol{U}_{\mathrm{b}}^{0},\boldsymbol{U}_{\mathrm{a}}^{0})$ is mistaken for another pair $(\boldsymbol{U}_{\mathrm{b}},\boldsymbol{U}_{\mathrm{a}})$ can be bounded by

$$\begin{aligned} &E\left[P(\boldsymbol{U}_{\mathrm{b}}^{0},\boldsymbol{U}_{\mathrm{a}}^{0}\rightarrow\boldsymbol{U}_{\mathrm{b}},\boldsymbol{U}_{\mathrm{a}})\right] \\ &\quad\leq\left|\boldsymbol{I}-\frac{1}{2}(\boldsymbol{U}_{\mathrm{b}}\boldsymbol{U}_{\mathrm{a}}\boldsymbol{U}_{\mathrm{a}}^{0H}\boldsymbol{U}_{\mathrm{b}}^{0H}+\boldsymbol{U}_{\mathrm{b}}^{0}\boldsymbol{U}_{\mathrm{a}}^{0}\boldsymbol{U}_{\mathrm{a}}^{H}\boldsymbol{U}_{\mathrm{b}}^{H})\right|^{-n_r}\cdot\left(\frac{\rho^{2}}{4\sigma^{2}}\right)^{-n_r n_t} \end{aligned} \tag{4.33}$$

A bound such as (4.31) or (4.33) was used in, for instance, Hughes et al. [18] for the design of (uniform) differential space-time codes. However, although it may be thought of as a feasible approach, an attempt to minimize this bound in the context of *nonuniform* space-time modulation may *not* produce the desired result since the target error rates for $\boldsymbol{U}_{\mathrm{b}}$ and $\boldsymbol{U}_{\mathrm{a}}$ are different.

4.3.1 Design Criteria

Since the additional message can be decoded only at high SNR, the matrices $\{\boldsymbol{U}_{\mathrm{a}}\}$ associated with the additional message should approximately be close to the identity matrix. Inspired by this observation, we can first consider the design of $\{\boldsymbol{U}_{\mathrm{b}}\}$, treating the presence of $\boldsymbol{U}_{\mathrm{a}}$ as an unmodelled noise-like disturbance term. Doing so, we can approximately bound the error probability for $\boldsymbol{U}_{\mathrm{b}}$ alone (assuming differential detection) by

$$E\left[P(\boldsymbol{U}_{\mathrm{b}}^{0}\rightarrow\boldsymbol{U}_{\mathrm{b}})\right]\lesssim\left|\boldsymbol{I}-\frac{1}{2}\left(\boldsymbol{U}_{\mathrm{b}}^{(0)}\boldsymbol{U}_{\mathrm{b}}^{H}+\boldsymbol{U}_{\mathrm{b}}\boldsymbol{U}_{\mathrm{b}}^{(0)H}\right)\right|^{-n_r}\cdot\left(\frac{\rho^{2}}{4\sigma^{2}+\alpha^{2}}\right)^{-n_r n_t} \tag{4.34}$$

where α^2 is a factor that incorporates the noise-like effect of the presence of $\boldsymbol{U}_{\mathrm{a}}$. Although the "bound" (4.34) is somewhat heuristic (and probably neither tight nor very accurate), we believe that it may serve a purpose as a meaningful design criterion for $\{\boldsymbol{U}_{\mathrm{b}}\}$.

Next, for the design of $\{\boldsymbol{U}_{\mathrm{a}}\}$ we proceed as follows. Suppose that the SNR is in a region such that $\boldsymbol{U}_{\mathrm{b}}$ can be reliably decoded. For the design of $\{\boldsymbol{U}_{\mathrm{a}}\}$ this

should be a reasonable assumption, since if it is not true then decoding of $\boldsymbol{U}_{\mathrm{a}}$ is probably of less interest anyway. Assuming that $\boldsymbol{U}_{\mathrm{b}}$ is known, the demodulation of $\boldsymbol{U}_{\mathrm{a}}$ is essentially another noncoherent detection problem. To obtain a criteria for the design of the constellation $\{\boldsymbol{U}_{\mathrm{a}}\}$, we want to form an error bound on $\boldsymbol{U}_{\mathrm{a}}$ and average it over all possible basic messages $\{\boldsymbol{U}_{\mathrm{b}}\}$. Using the bound (4.25) for noncoherent detection appears to yield criteria that are very difficult to use. As a suboptimal approach we used instead the bound (4.24) for coherent detection. Doing so results in the following criterion

$$\begin{aligned} E\left[P(\boldsymbol{U}_{\mathrm{a}}^{0} \to \boldsymbol{U}_{\mathrm{a}})\right] \leq & \frac{1}{|\mathcal{X}_{\mathrm{b}}|} \sum_{\boldsymbol{U}_{\mathrm{b}} \in \mathcal{X}_{\mathrm{b}}} \left|(\boldsymbol{U}_{\mathrm{b}}\boldsymbol{U}_{\mathrm{a}}^{0} - \boldsymbol{U}_{\mathrm{b}}\boldsymbol{U}_{\mathrm{a}})(\boldsymbol{U}_{\mathrm{b}}\boldsymbol{U}_{\mathrm{a}}^{0} - \boldsymbol{U}_{\mathrm{b}}\boldsymbol{U}_{\mathrm{a}})^{H}\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \\ = & \left|(\boldsymbol{U}_{\mathrm{a}}^{0} - \boldsymbol{U}_{\mathrm{a}})(\boldsymbol{U}_{\mathrm{a}}^{0} - \boldsymbol{U}_{\mathrm{a}})^{H}\right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \end{aligned} \tag{4.35}$$

where in the last step $\boldsymbol{U}_{\mathrm{b}}$ disappears since it is unitary. In (4.35), $|\cdot|$ denotes the number of elements of the set.

4.3.2 Union Bound on the Error Probability

By the union bound, the probability for the basic message to be in error (averaged over all possible pairs of additional messages) can be bounded by

$$\begin{aligned} & E\left[P(\text{basic message in error})\right] \\ \leq & \frac{1}{|\mathcal{X}_{\mathrm{b}}|} \sum_{\substack{(\boldsymbol{U}_{\mathrm{b}}^{(k)}, \boldsymbol{U}_{\mathrm{b}}^{(n)}) \in \mathcal{X}_{\mathrm{b}} \\ k \neq n}} \frac{1}{|\mathcal{X}_{\mathrm{a}}|^2} \sum_{(\boldsymbol{U}_{\mathrm{a}}^{(r)}, \boldsymbol{U}_{\mathrm{a}}^{(s)}) \in \mathcal{X}_{\mathrm{a}}} \\ & \left| \boldsymbol{I} - \frac{1}{2}\left(\boldsymbol{U}_{\mathrm{b}}^{(k)}\boldsymbol{U}_{\mathrm{a}}^{(r)}\boldsymbol{U}_{\mathrm{a}}^{(s)H}\boldsymbol{U}_{\mathrm{b}}^{(n)H} + \boldsymbol{U}_{\mathrm{b}}^{(n)}\boldsymbol{U}_{\mathrm{a}}^{(s)}\boldsymbol{U}_{\mathrm{a}}^{(r)H}\boldsymbol{U}_{\mathrm{b}}^{(k)H}\right) \right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2}\right)^{-n_r n_t} \end{aligned} \tag{4.36}$$

Likewise, the error rate for the additional message (averaged over all possible pairs of basic messages), can be bounded by a similar expression

$$
\begin{aligned}
&E\,[P(\text{additional message in error})] \\
\leq &\frac{1}{|\mathcal{X}_{\mathrm{a}}|} \sum_{\substack{(\boldsymbol{U}_{\mathrm{a}}^{(r)}, \boldsymbol{U}_{\mathrm{a}}^{(s)}) \in \mathcal{X}_{\mathrm{a}} \\ r \neq s}} \frac{1}{|\mathcal{X}_{\mathrm{b}}|^2} \sum_{(\boldsymbol{U}_{\mathrm{b}}^{(k)}, \boldsymbol{U}_{\mathrm{b}}^{(n)}) \in \mathcal{X}_{\mathrm{b}}} \\
&\left| \boldsymbol{I} - \frac{1}{2}\left(\boldsymbol{U}_{\mathrm{b}}^{(k)} \boldsymbol{U}_{\mathrm{a}}^{(r)} \boldsymbol{U}_{\mathrm{a}}^{(s)H} \boldsymbol{U}_{\mathrm{b}}^{(n)H} + \boldsymbol{U}_{\mathrm{b}}^{(n)} \boldsymbol{U}_{\mathrm{a}}^{(s)} \boldsymbol{U}_{\mathrm{a}}^{(r)H} \boldsymbol{U}_{\mathrm{b}}^{(k)H} \right) \right|^{-n_r} \cdot \left(\frac{\rho^2}{4\sigma^2} \right)^{-n_r n_t}
\end{aligned}
\tag{4.37}
$$

These bounds will be useful for performance optimization.

4.4 Design Examples

4.4.1 $R_{\mathrm{b}} = 1$, $R_{\mathrm{a}} = 1$ Code for 2 TX Antennas

We first construct a code where the rate for the basic message is $R_{\mathrm{b}} = 1$ bit/sec/Hz and the rate for the additional message is $R_{\mathrm{a}} = 1$ bit/sec/Hz as well. We take the basic message $\boldsymbol{U}_{\mathrm{b}}$ taken from the following set

$$
\mathcal{X}_{\mathrm{b}} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\} \tag{4.38}
$$

The constellation in (4.38), which is uniform and possesses certain optimality properties, is due to Hughes et al. [18] and was called "BPSK" therein. Based on the design rules in Section 4.3.1, we have handcrafted the following constellation of matrices for the additional message $\boldsymbol{U}_{\mathrm{a}}$

$$
\mathcal{X}_{\mathrm{a}} = \left\{ \begin{bmatrix} e^{i\pi\lambda} & 0 \\ 0 & e^{i\pi\gamma} \end{bmatrix}, \begin{bmatrix} e^{i\pi\gamma} & 0 \\ 0 & e^{i\pi\lambda} \end{bmatrix}, \begin{bmatrix} e^{-i\pi\lambda} & 0 \\ 0 & e^{-i\pi\gamma} \end{bmatrix}, \begin{bmatrix} e^{-i\pi\gamma} & 0 \\ 0 & e^{-i\pi\lambda} \end{bmatrix} \right\} \tag{4.39}
$$

where (λ, γ) are design parameters (to be discussed below).

The constellation may appear to be somewhat arbitrary but it possesses some nice features. First, all matrices in (4.39) are unitary, this is a requirement for MIMO differential encoding. Second, it is symmetric and uniform which means that the error probability performance should be the same for all constellation points. Also similarly to the nonuniform PSK encoding for a single channel (cf. Section ??), multiplication with the matrices in (4.39) can be interpreted as adding a small phase shift to the elements of the basic message $\boldsymbol{U}_{\mathrm{b}}$.

We may also think of other forms of matrices as the additional message. For instance, consider the following anti-diagonal matrix

$$\boldsymbol{U}_{\mathrm{a}} = \begin{bmatrix} 0 & e^{i\pi\lambda} \\ e^{i\pi\gamma} & 0 \end{bmatrix}. \tag{4.40}$$

But we immediately find that this does not work. Consider, for example,

$$\boldsymbol{U}_{\mathrm{b}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.41}$$

Then,

$$\boldsymbol{U}_{\mathrm{b}} \cdot \boldsymbol{U}_{\mathrm{a}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & e^{i\pi\lambda} \\ e^{i\pi\gamma} & 0 \end{bmatrix} = \begin{bmatrix} 0 & e^{i\pi\lambda} \\ e^{i\pi\gamma} & 0 \end{bmatrix} \tag{4.42}$$

It is seen that multiplying this additional message $\boldsymbol{U}_{\mathrm{a}}$ not only introduces a small phase shift but also interchanges the columns of the original basic message $\boldsymbol{U}_{\mathrm{b}}$. This is not acceptable since our design goal is to have the additional message only pose a small disturbance to the basic message so that for those less-capable receivers can still detect the basic message even without knowledge of the existence of the additional message. The $\boldsymbol{U}_{\mathrm{a}}$ here simply destroys the original basic message constellation so it is not feasible.

Let us now return to our design task. If we take the initial transmit matrix, somewhat arbitrarily

$$\boldsymbol{X}(0) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \tag{4.43}$$

The factor $\frac{1}{\sqrt{2}}$ is inserted to normalize the transmitted power to unity for each time interval. It follows that

$$\boldsymbol{X}^H(t)\boldsymbol{X}(t) = \boldsymbol{I} \tag{4.44}$$

for all t and hence the differential encoding of the code is meaningful. Furthermore, we can verify that all elements of $\boldsymbol{X}$ have always constant magnitude:

$$\left|X_{k,l}(t)\right| = \frac{1}{\sqrt{2}} \tag{4.45}$$

and hence all transmitted symbols are obtained by constant envelope modulation, which was one of our design goals.

The error performance (of both the basic and the additional message) will depend on λ and γ, and typically the error performance of the additional message can be traded against that of the basic message. The values (λ,γ) must be chosen carefully. For instance, at first glance intuition would perhaps suggest us to take $\lambda > 0$ and $\gamma = 0$, but we can show that such a "simple" choice leads to a code that "works" but that does *not* provide maximal diversity.

In our experiments, we used the union bound (4.36) and (4.37) to optimize over (λ, γ). In particular, for a given "tolerable" loss in performance for the basic message, we can find the pair (λ, γ) that minimizes (4.37). The result is shown in Figure 4–5. For example, if we can accept a degradation of 1.5 dB for the basic message, then $\lambda = 0.2$ and $\gamma = 0.035$ are optimal. The optimization over (λ, γ) is

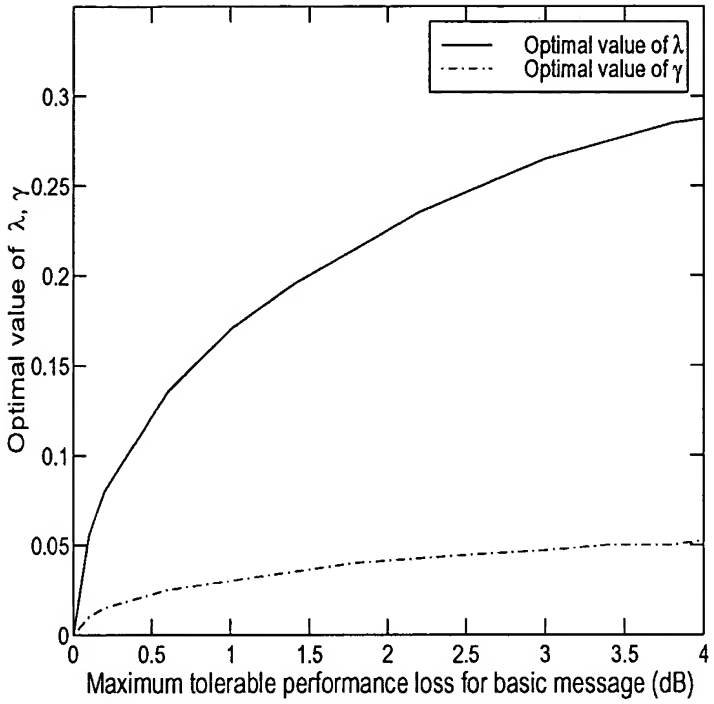

Figure 4–5: Optimal values of (λ, γ) for the $R_b = 1$, $R_a = 1$ code, for a given acceptable performance degradation of the basic message. The results are obtained via minimization of the union bound (4.36) and a corresponding expression for the error rate of the additional message.

further illustrated in Figure 4–6, where we show how the performance associated with the basic and additional messages varies with (λ, γ).

Figure 4–7 shows the empirical bit-error-rate (BER), obtained via Monte-Carlo simulation, for the code described above using $\lambda = 0.2$, $\gamma = 0.035$ and ML decoding. The solid lines ("—") show the performance of differential nonuniform BPSK for a conventional system with $n_t = 1$ transmit antenna and a single receive antenna (this is essentially a special case of Pursley et al. [8]). The dashed lines ("- -") show the performance for a system with $n_t = 2$ (and a single receive antenna) using the new code presented above. For the curves without marks, only a basic message is transmitted. The curves with marks show the performance when both a basic and an additional message are transmitted: the

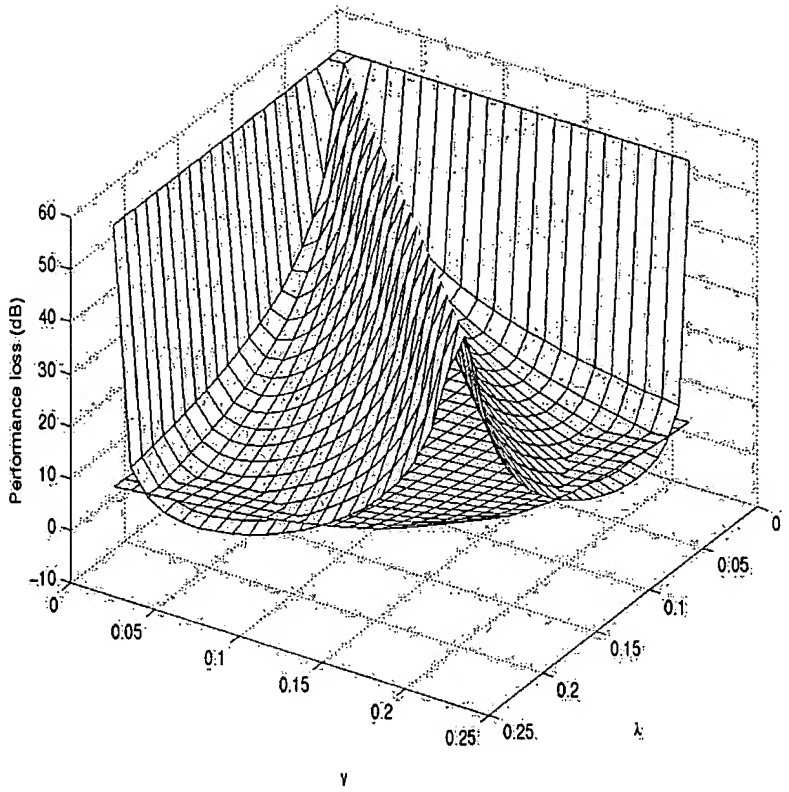

Figure 4–6: Performance degradation for the basic message and performance gain for the additional message for the $R_b = 1$, $R_a = 1$ code. The results are obtained via (4.36), along with a corresponding expression for the additional message. The curves for the basic message are normalized relative to the "undisturbed case" ($\lambda = \gamma = 0$) and the curves for the additional message are normalized relative to $\lambda = 0.2$, $\gamma = 0.035$.

curves marked with "o" show the BER for the basic message, and the curves with "×" show the BER for the additional message. Clearly, the transmit diversity system outperforms the conventional one – observe, in particular, the different slopes of the BER curves both for the basic and for the additional message. The simulation also confirms that the transmission of an additional message incurs a small performance degradation for the basic message.

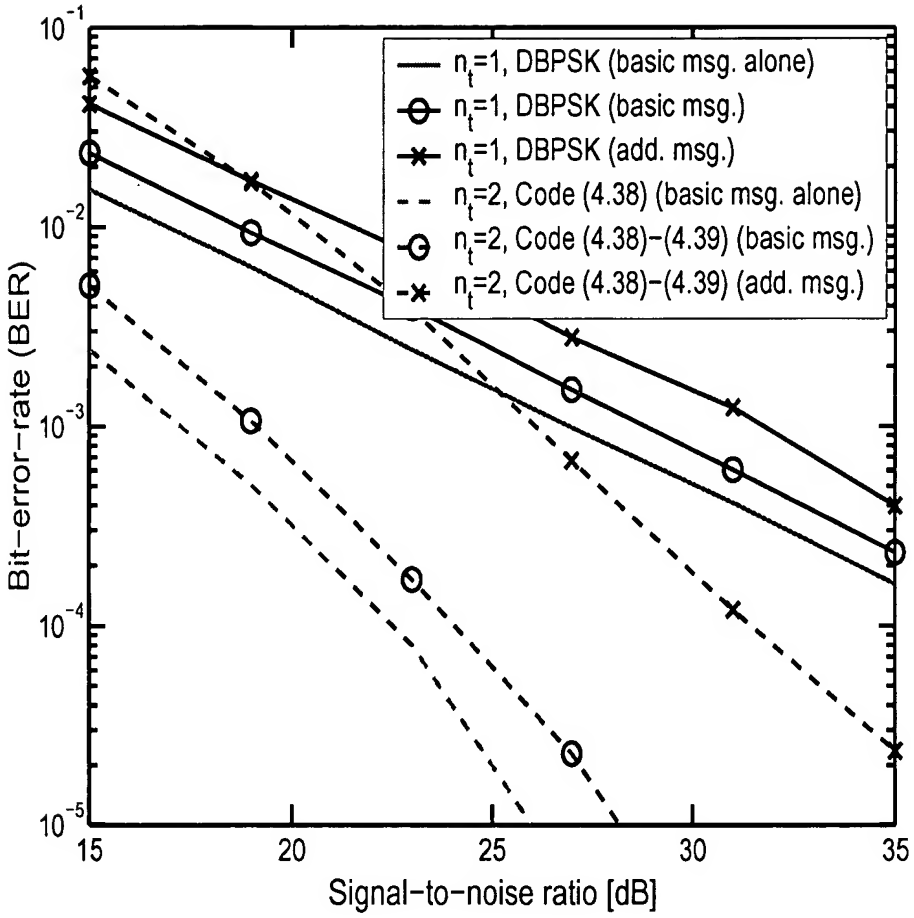

Figure 4–7: Empirical BER for the $R_b = 1$, $R_a = 1$ code with $\lambda = 0.2$, $\gamma = 0.035$.

4.4.2 $R_b = 2$, $R_a = 1$ Code

To obtain a code with a higher information rate for the basic message, we next take $\boldsymbol{U}_b$ from the following algebraic group of 16 matrices (which is due to Hughes et al. [18] as well) generated by

$$\mathcal{X}_b = \left\langle \begin{bmatrix} e^{j\pi/4} & 0 \\ 0 & e^{-j\pi/4} \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \right\rangle \tag{4.46}$$

The notation in (4.46) means that all possible matrices can be generated by choosing any arbitrary integers M and N in the following expression

$$\begin{bmatrix} e^{j\pi/4} & 0 \\ 0 & e^{-j\pi/4} \end{bmatrix}^M \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^N \tag{4.47}$$

The constellation used for the additional message is chosen to be the same as (4.39), so we now have a system with rate $R_b = 2$ for the basic message and rate $R_a = 1$ for the additional message. The corresponding simulated BER is shown in Figure 4–8.

4.4.3 Receive Diversity

To demonstrate the transmit diversity achieved for the 2-TX system in the simplest way, our results above are based on the 1-RX case. Although of this, our newly proposed nonuniform STBC codes are not only useful for 1-RX case but also for a general n_r-RX case. This is obvious since our theoretical derivations above are not tailored for the 1-RX case, hence it should work automatically for a general n_r-RX case. Also we could expect a diversity order of $n_r n_t$ can be achieved. This is so since we can see the data model of a n_r-RX n_t-TX system as

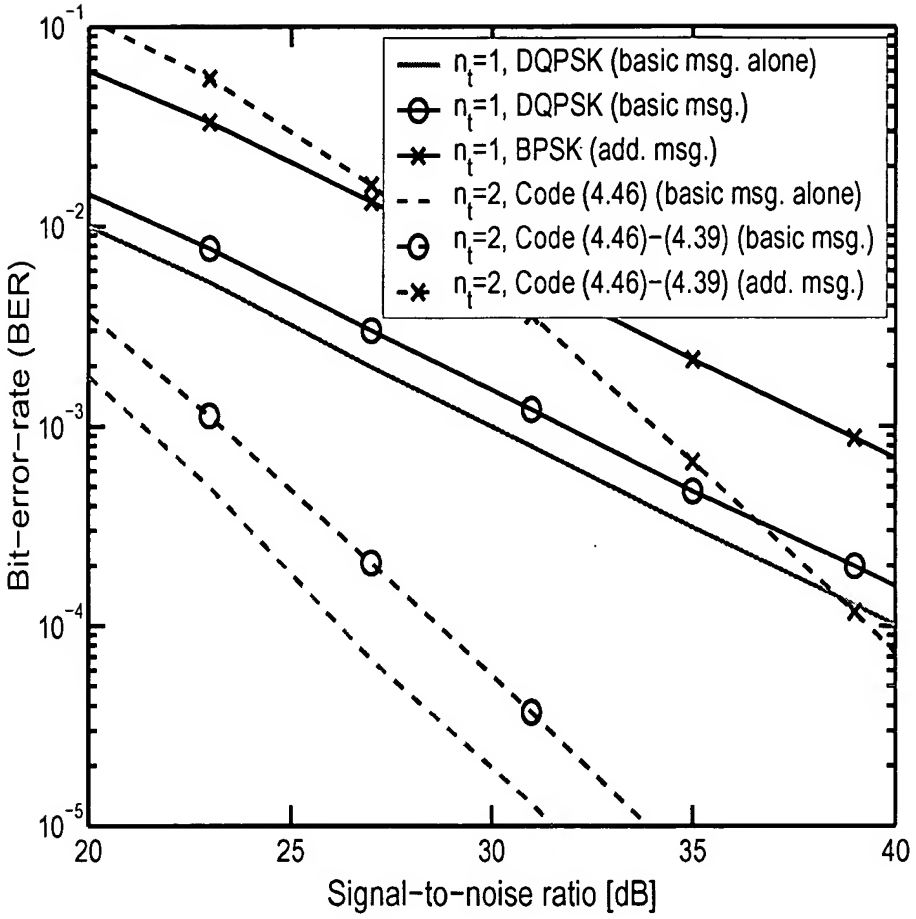

Figure 4–8: Empirical BER for the $R_b = 2$, $R_a = 1$ nonuniform space-time constellation. The parameters were $\lambda = 0.078$, $\gamma = 0.018$ (found via optimization of (4.36)).

n_r different 1-RX n_t-TX system

$$
\begin{aligned}
& \boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E} \\
\Longleftrightarrow\ & \begin{bmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_{n_r} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_{n_r} \end{bmatrix} \boldsymbol{X} + \begin{bmatrix} \boldsymbol{E}_1 \\ \vdots \\ \boldsymbol{E}_{n_r} \end{bmatrix} \\
\Longleftrightarrow\ & \begin{cases} \boldsymbol{Y}_1 & = \boldsymbol{H}_1\boldsymbol{X} + \boldsymbol{E}_1 \\ \vdots & \\ \boldsymbol{Y}_{n_r} & = \boldsymbol{H}_{n_r}\boldsymbol{X} + \boldsymbol{E}_{n_r} \end{cases}
\end{aligned}
\tag{4.48}
$$

where $\boldsymbol{Y}_k$ for $k = 1, \cdots, n_r$ is the $1 \times N$ received data at the kth receive antenna, $\boldsymbol{H}_k$ is the $1 \times n_t$ channel matrix associated with the kth receive antenna, $\boldsymbol{E}_k$ is the $1 \times n_t$ noise matrix at the kth receive antenna.

As usual, for simplicity if we assume the channel matrices associated with n_r different receive antennas $\boldsymbol{H}_k$ (for $k = 1, \cdots, n_r$) are independent and the noise is white ($\boldsymbol{E}_1, \cdots, \boldsymbol{E}_{n_r}$ are uncorrelated), then the n_r-RX n_t-TX can be broken down into n_r number of independent 1-RX n_t-TX system. Each receive antenna has sufficient data to detect the transmitted matrix $\boldsymbol{X}$ so they can work individually and achieve transmit diversity of order n_t. To further utilize receive diversity we can employ joint ML detection across all n_r independent receive antennas thus increasing the overall diversity by a factor of n_r. Since the n_r receive antennas are independent at the receiver site, so optimizing the space-time encoding scheme for a single receive antenna is equivalent to optimizing a system with n_r independent receive antennas.

4.5 Suboptimal Detector for Nonuniform D-STBC

To detect both the basic and the additional messages in a joint ML fashion we have to search through all $|\mathcal{X}_\mathrm{b}| \cdot |\mathcal{X}_\mathrm{a}|$ possible combinations to find the combination of $\boldsymbol{U}_\mathrm{b}$ and $\boldsymbol{U}_\mathrm{a}$ which maximize the cost function in (4.30). This may be computationally burdensome if the constellation sizes of both layers are large. A possible remedy to this problem is to use a suboptimal detection rule described as follows

Step 1. Detect the basic message only using (4.17), neglecting the presence of the additional message

$$\begin{aligned} \hat{\boldsymbol{U}}_\mathrm{b}(t) &= \min_{\boldsymbol{U}_\mathrm{b}(t) \in \mathcal{X}_\mathrm{b}} \left\| \begin{bmatrix} \boldsymbol{Y}(t-1) & \boldsymbol{Y}(t) \end{bmatrix} \cdot \Pi^{\perp}_{[I \ \boldsymbol{U}_\mathrm{b}(t)]^H} \right\|^2 \\ &\Longleftrightarrow \max_{\boldsymbol{U}_\mathrm{b}(t) \in \mathcal{X}_\mathrm{b}} \mathrm{Re}\left\{ \mathrm{Tr}\left\{ \boldsymbol{U}_\mathrm{b}(t) \boldsymbol{Y}^H(t) \boldsymbol{Y}(t-1) \right\} \right\} \end{aligned} \tag{4.49}$$

Step 2. Detect the additional message assuming the decision taken in *Step 1* is correct

$$\begin{aligned} &\min_{\boldsymbol{U}_{\mathrm{a}} \in \mathcal{X}_{\mathrm{a}}} \left\| \left[\boldsymbol{Y}(t-1) \cdot \hat{\boldsymbol{U}}_{\mathrm{b}}(t) \quad \boldsymbol{Y}(t) \right] \cdot \Pi^{\perp}_{[I \ \boldsymbol{U}_{\mathrm{a}}]^H} \right\|^2 \\ &\Longleftrightarrow \max_{\boldsymbol{U}_{\mathrm{a}}(t) \in \mathcal{X}_{\mathrm{a}}} \mathrm{Re} \left\{ \mathrm{Tr} \left\{ \boldsymbol{U}_{\mathrm{a}}(t) \boldsymbol{Y}^H(t) \boldsymbol{Y}(t-1) \hat{\boldsymbol{U}}_{\mathrm{b}}(t) \right\} \right\} \end{aligned} \tag{4.50}$$

Let us elaborate more on why this suboptimal detection scheme works. Since $\boldsymbol{U}_{\mathrm{a}}$ is close to the identity matrix, when compared to the basic message $\boldsymbol{U}_{\mathrm{b}}$ it can be treated as small unmodelled noise; hence the detection rule in *Step 1* can still be able to detect the basic message with only a small performance degradation. As the error performance of the basic message is typically superior to that of the additional message, we can assume the detected basic message is correct when we are dealing with the detection of the additional message. Thus the observation model for the detection of $\boldsymbol{U}_{\mathrm{a}}(t)$ becomes

$$\begin{aligned} &\left[\boldsymbol{Y}(t-1) \cdot \hat{\boldsymbol{U}}_{\mathrm{b}}(t) \quad \boldsymbol{Y}(t) \right] \\ &\quad = \boldsymbol{H}\boldsymbol{X}(t-1) \cdot \left[\hat{\boldsymbol{U}}_{\mathrm{b}}(t) \quad \hat{\boldsymbol{U}}_{\mathrm{b}}(t) \boldsymbol{U}_{\mathrm{a}}(t) \right] + \left[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t) \right] \\ &\quad = \boldsymbol{H}\boldsymbol{X}(t-1) \hat{\boldsymbol{U}}_{\mathrm{b}}(t) \cdot \left[\boldsymbol{I} \quad \boldsymbol{U}_{\mathrm{a}}(t) \right] + \left[\boldsymbol{E}(t-1) \quad \boldsymbol{E}(t) \right] \end{aligned} \tag{4.51}$$

Treating $\boldsymbol{H}\boldsymbol{X}(t-1)\hat{\boldsymbol{U}}_{\mathrm{b}}(t)$ as an unknown "effective" channel as in (4.29) and using (4.17) and (4.30) yields the detection rule in *Step 2.*

By applying this suboptimal detection rule the computational complexity is reduced from $|\mathcal{X}_{\mathrm{b}}| \cdot |\mathcal{X}_{\mathrm{a}}|$ to $|\mathcal{X}_{\mathrm{b}}| + |\mathcal{X}_{\mathrm{a}}|$. For example, for the $R_{\mathrm{b}} = 2$, $R_{\mathrm{a}} = 1$ code, the computational time is reduced from $|\mathcal{X}_{\mathrm{b}}| \cdot |\mathcal{X}_{\mathrm{a}}| = 16 \cdot 4 = 64$ to $|\mathcal{X}_{\mathrm{b}}| + |\mathcal{X}_{\mathrm{a}}| = 16 + 4 = 20$. For this example the reduction is not very significant; yet this suboptimal detection scheme does provide more flexibility as opposed

to the joint ML detection because if only the *basic* message is needed, the computational load will be further reduced to $|\mathcal{X}_b| = 16$ (for this example).

We expect that there is some performance loss associated with this suboptimal detection scheme since it does not exploit knowledge of the structure of the additional message in the *Step 1*.

The Monte-Carlo simulation result in Figure 4–9 shows the performance of the suboptimal detector for the $R_b = 2$, $R_a = 1$ code compared to the performance of the joint ML detector. The detection of the basic message gives a small performance loss (≈ 0.3 dB) compared to the joint ML detector, whereas the detection of the additional message is (to within practical accuracy) equal to that of the joint ML detector.

4.6 Comparison of Differential Group Codes and A Differential Alamouti Code

4.6.1 Differential Alamouti Code

The main advantage of the nonuniform space-time group codes introduced in Section 4.4 is that the signal envelope is constant. However, also for this reason they cannot achieve the maximum possible performance (in terms of BER) because they do not make use of any amplitude information in the signal. On the contrary, orthogonal space-time block codes (OSTBC) should provide better performance (BER) since they are already optimized by construction [17, 23, 14]. For 2 TX antennas OSTBC reduces to Alamouti code. The construction of an Alamouti code matrix is as follows

$$\boldsymbol{U}(t) = \frac{1}{\sqrt{2}} \begin{bmatrix} s_1(t) & s_2^*(t) \\ s_2(t) & -s_1^*(t) \end{bmatrix} \tag{4.52}$$

where s_1 and s_2 are the information bearing symbols and $(\cdot)^*$ stands for the complex conjugate. For the $R_b = 1$ system, s_1 and s_2 are taken from the BPSK set $\{-1, 1\}$, and for the $R_b = 2$ system, they are taken from the QPSK set

$\{-1, -j, 1, j\}$. The $1/\sqrt{2}$ factor is included for normalization purposes so that the transmitted power per time interval equals one. This Alamouti matrix $\boldsymbol{U}(t)$ can then be encoded differentially by using (4.28) in Section 4.2. Note that the Alamouti code works together with the additional message $\boldsymbol{U}_{\mathrm{a}}$ taken from (4.39) to form a nonuniform constellation, and differentially encoded by (4.32), but unfortunately the matrices in the so-obtained constellation do not have constant signal envelope.

Here we compare the performance of the group codes to that of the Alamouti code. The $R_{\mathrm{b}} = 1$ group codes (4.38) in Section 4.4 and $R_b = 1$ Alamouti codes are indeed equivalent, so there is no performance advantage for Alamouti codes for this case. The equivalence of both codes is easy to show as in this case the set of Alamouti matrices become

$$\mathcal{X}_{\text{alamouti}} = \left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \right\} \quad (4.53)$$

If we pre-multiply the above set by $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ then we obtain the $R_b = 1$ group code in (4.38).

For the $R_b = 2$ codes, the Alamouti code does provide somewhat better BER performance than the group codes in (4.46) and (4.39). This is so since the Alamouti code uses both signal amplitude and phase to transmit information; thus it should be less susceptible to the phase disturbance caused by the additional message. Figure 4–10 shows the simulated performance difference between the $R_{\mathrm{b}} = 2$, $R_{\mathrm{a}} = 1$ nonuniform group codes in (4.46) and (4.39) and a $R_{\mathrm{b}} = 2$ Alamouti code together with the additional message (4.39). For the basic message the Alamouti code outperforms the group code by about 1dB, and it is somewhat more robust when additional message is added. Meanwhile the

performances of the detection of the additional message are equal both when the Alamouti and the group codes (4.39) are used as the basic message carrier.

4.6.2 Alamouti Code with Differentially Encoded Symbols

It may be argued that both the constant signaling envelope and maximum BER performance can be achieved at the same time by using symbol-wise differential encoding and forming the Alamouti matrix afterwards. However, unfortunately this encoding scheme does not work. This is demonstrated as follows.

Suppose that we employ symbol-wise differential encoding

$$\begin{aligned} x_1(t) &= x_1(t-1) \cdot s_1(t) \\ x_2(t) &= x_2(t-1) \cdot s_2(t) \end{aligned} \tag{4.54}$$

and use these elements to form the Alamouti matrix

$$\begin{aligned} \boldsymbol{X}(t-1) &= \begin{bmatrix} x_1(t-1) & x_2^*(t-1) \\ x_2(t-1) & -x_1^*(t-1) \end{bmatrix} \\ \boldsymbol{X}(t) &= \begin{bmatrix} x_1(t) & x_2^*(t) \\ x_2(t) & -x_1^*(t) \end{bmatrix} = \begin{bmatrix} x_1(t-1)s_1(t) & x_2^*(t-1)s_2^*(t) \\ x_2(t-1)s_2(t) & -x_1^*(t-1)s_1^*(t) \end{bmatrix} \end{aligned} \tag{4.55}$$

The above code matrix has the property that $|x_1(t)| = |x_2(t)| = |s_1(t)| = |s_2(t)| = 1$ for all t and hence the signal envelope is constant for both transmit antennas all the time. Also $\boldsymbol{X}(t)$ is unitary for all t since it is an Alamouti matrix. The observation model (4.29) becomes (cf. Section 4.2)

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{Y}(t-1) & \boldsymbol{Y}(t) \end{bmatrix} \\ =&\boldsymbol{H}\begin{bmatrix} \boldsymbol{X}(t-1) & \boldsymbol{X}(t) \end{bmatrix} + \begin{bmatrix} \boldsymbol{E}(t-1) & \boldsymbol{E}(t) \end{bmatrix} \\ =&\boldsymbol{H}\boldsymbol{X}(t-1)\begin{bmatrix} \boldsymbol{I} & \boldsymbol{X}^H(t-1)\boldsymbol{X}(t) \end{bmatrix} + \begin{bmatrix} \boldsymbol{E}(t-1) & \boldsymbol{E}(t) \end{bmatrix} \end{aligned} \tag{4.56}$$

Employing the noncoherent detection rule (4.30 to the above model (4.56) the following detection rule is obtained

$$\begin{aligned}
&\max_{s_1(t),s_2(t)\in\mathcal{S}} \ \mathrm{Re}\Big\{ \ \mathrm{Tr}\Big\{(\boldsymbol{X}^H(t-1)\boldsymbol{X}(t)\boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1))\Big\} \ \Big\} \\
\Longleftrightarrow &\max_{s_1(t),s_2(t)\in\mathcal{S}} \\
&\mathrm{Re}\left\{ \ \mathrm{Tr}\left\{ \begin{bmatrix} x_1^*(t-1) & x_2^*(t-1) \\ x_2(t-1) & -x_1(t-1) \end{bmatrix} \begin{bmatrix} x_1(t-1)\cdot s_1(t) & x_2^*(t-1)\cdot s_2^*(t) \\ x_2(t-1)\cdot s_2(t) & -x_1^*(t-1)\cdot s_1^*(t) \end{bmatrix} \boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1) \right\} \ \right\} \\
\Longleftrightarrow &\max_{s_1(t),s_2(t)\in\mathcal{S}} \\
&\mathrm{Re}\left\{ \ \mathrm{Tr}\left\{ \begin{bmatrix} s_1(t)+s_2(t) & x_1^*(t-1)x_2^*(t-1)\cdot(s_2^*(t)-s_1^*(t)) \\ x_1(t-1)x_2(t-1)\cdot(s_1(t)-s_2(t)) & s_1^*(t)+s_2^*(t) \end{bmatrix} \boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1) \right\} \right\}
\end{aligned} \tag{4.57}$$

The above maximization problem (4.57) cannot be solved unless $x_1(t-1)$ and $x_2(t-1)$ are known. For example, consider the following two sets of values

$$\{s_1(t), s_2(t), x_1(t-1)x_2(t-1)\} = \{1, -1, 1\} \tag{4.58}$$

and

$$\{s_1(t), s_2(t), x_1(t-1)x_2(t-1)\} = \{-1, 1, -1\} \tag{4.59}$$

These make the metric in (4.57) equal to

$$\mathrm{Re}\left\{ \ \mathrm{Tr}\left\{ \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix} \boldsymbol{Y}^H(t)\boldsymbol{Y}(t-1) \right\} \ \right\} \tag{4.60}$$

which implies that the sets $\{s_1, s_2\} = \{1, -1\}$ and $\{s_1, s_2\} = \{-1, 1\}$ are not identifiable if $x_1(t-1)$ and $x_2(t-1)$ are unknown.

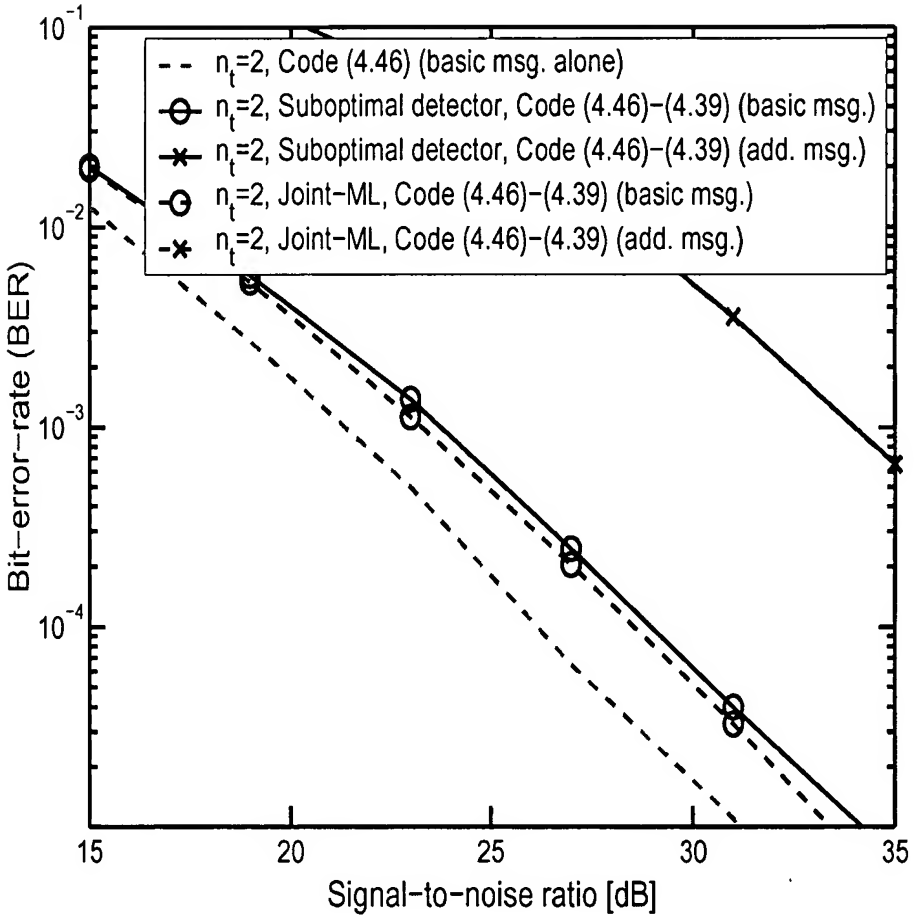

Figure 4-9: Monte-Carlo Simulation of the BER performance for the suboptimal receiver (4.49) and (4.50) in Section 4.5, for the joint ML receiver and for the basic message only system with $R_b = 2$, $R_a = 1$ code. The curves "$-\ \times\ -$" and "$-\ \times\ -$" overlap.

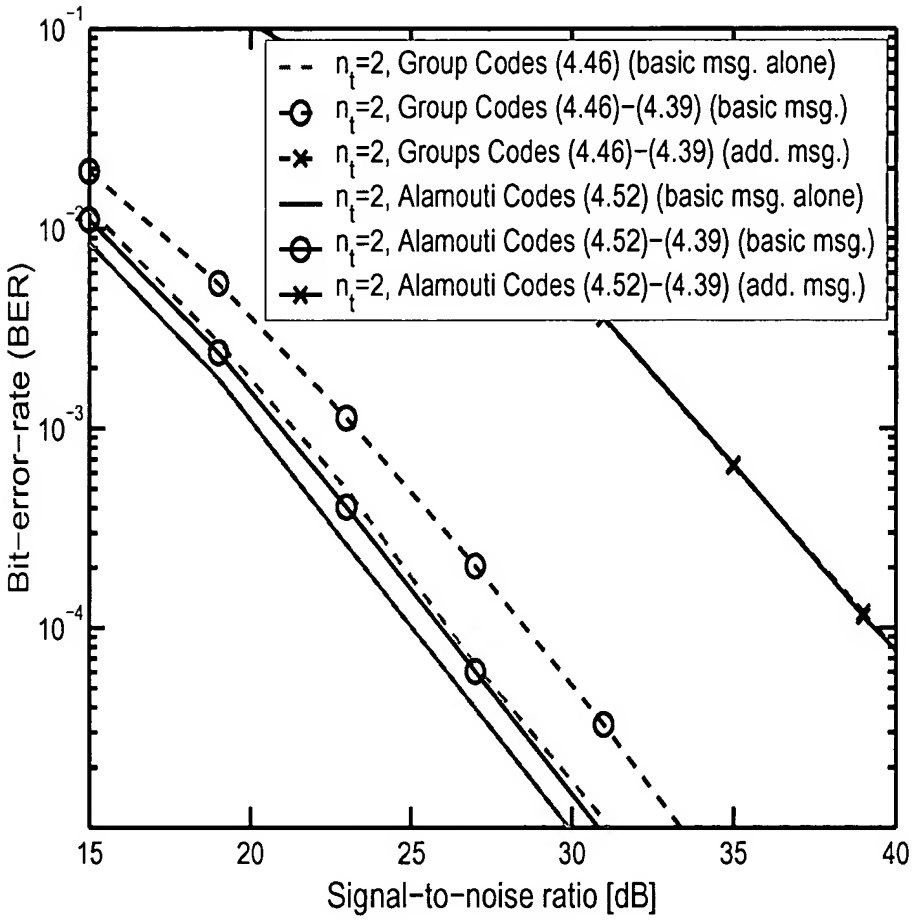

Figure 4–10: Monte Carlo Simulation result for the BER performance of $R_b = 2$, $R_a = 1$ system, with differential group codes and differential Alamouti codes. Note that the solid and dash curves with "×" marks overlap.

CHAPTER 5
CONVOLUTIONAL PRE-CODING

5.1 Introduction to Convolutional Codes

In Pursley et al. [12], convolutional coding are introduced for extra error protection before the data symbols are further encoded by the nonuniform PSK modulation. For completeness we introduce here the basic concept of convolutional coding as a tool to improve the BER performance.

Convolutional coding is a kind of forward error correction (FEC) technique, which means that no feedback channel is required in contrast to, e.g. Automatic Repeat Request (ARQ).

The idea of convolutional coding is simple: To improve the error performance of a transmission link by adding some carefully designed redundant bits to the data before it is transmitted through the channel. This process of adding redundancy is known as channel coding. In contrast to block codes which operate on large message blocks, convolutional codes operate on a continuous stream of data bits.

Convolutional coding is a FEC technique which is especially suitable to channels which are contaminated by additive white gaussian noise (AWGN). By using convolutional coding, BER performance can be improved significantly at the same SNR, at the cost of lowering the data rate. Meanwhile this rate loss can be compensated for by increasing the constellation size.

There are two important parameters for a particular convolutional codes, namely the *code rate* and the *constraint length*. The code rate, k/n, is simply the ratio of the number of data bits input to the convolutional encoder (k) to the

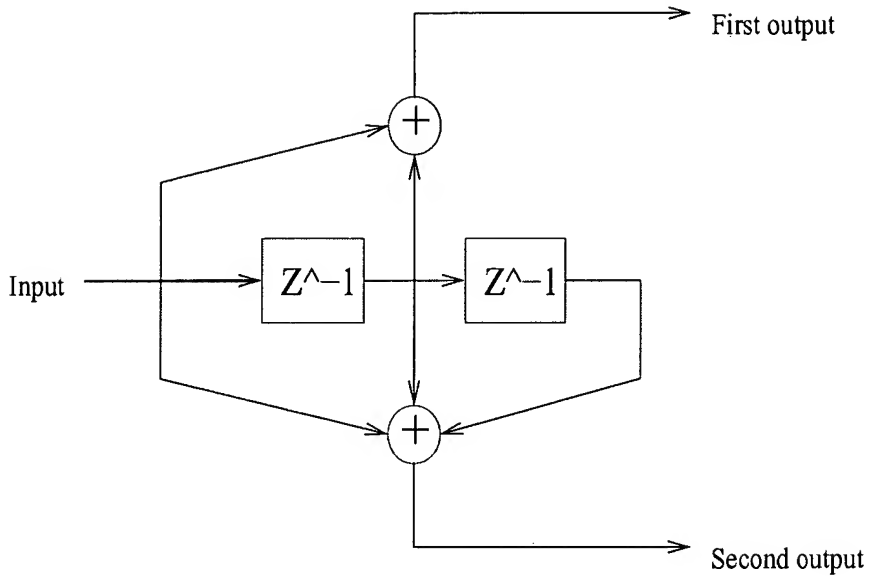

Figure 5–1: A simple rate 1/2 convolutional encoder.

number of output bits by the convolutional encoder (n), can also be a measure of the efficiency of the code. The constraint length, K, denotes the "length" or "memory" of the convolutional encoder, that is, each output bit is affected by K previous input bits.

For years, convolutional coding has been the predominant FEC technique. But there is a tradeoff of employing convolutional coding. For example, using a 1/2 rate convolutional codes, the data rate will be deducted to half, provided that the modulation technique is the same. This is simply because with rate 1/2 convolutional encoding, two channel symbols are needed to be transmitted per one data bit. But the advantage is that, the convolutional coding can improve the error performance at the same SNR, at the cost that the data rate will decrease by a factor of n/k.

Figure 5–1 shows a typical rate 1/2 convolutional encoder, built with two shift registers and two modulo-2 adders.

One should pay attention that an interleaver is usually used to interleave the encoded data bit stream from the output of the convolutional encoder. The benefit of interleaving the data bits is that it can provide *time diversity*: consecutive bits are likely to be contaminated by a temporary strong interference or noise, or sustained by deep fading channel environment, this can make the error correcting algorithm at the convolutional decoder fails to correct bit errors and recover the correct information. With sufficient interleaving, the originally consecutive bits are now separated far apart as they are being transmitted through the wireless channel, so that we can reasonably assume that the bits nearby each other after de-interleaving have undergone independent fading and are affected by independent noise and interference, this is desirable to the convolutional decoder as consecutive bits are not likely to encounter errors simultaneously. Also the design task of the ML detector at the receiver is also simplified.

5.2 Convolutional Codes Applied to Space-Time Coding System

We are going to investigate the error performance of the system if a convolutional encoder is employed to the information bit streams prior to STBC mapping. It is expected that the BER performance can be improved substantially by such kind of pre-coding in an adverse channel environment. This simply trades the error performance with data rate.

To start with, we choose a simple rate-1/2 convolutional encoder with constraint length $K = 3$, generators are $5, 7$ in octal.

As for our $R_b = 1, R_a = 1$ code, there are two bits carried in the basic layer and two bits carried in the additional layer for each transmit time interval, let s_1, s_2 denote the two bits carried in the basic layer and s_3, s_4 denote the twos bits carried in the additional layer, respectively. And now we have four separate

(independent) bit streams, for example they may look like the following

$$\begin{aligned} s_1 &= 0,1,1,0,1,1,0,1,1,\cdots \\ s_2 &= 1,1,0,1,1,0,0,0,1,\cdots \\ s_3 &= 0,0,0,1,0,1,1,1,1,\cdots \\ s_4 &= 0,0,1,0,0,1,0,1,0,\cdots \end{aligned} \tag{5.1}$$

We model the s_1, s_2, s_3, s_4 as independent bit streams with random 0 and 1. For each bit stream we pass it through a stand-alone binary convolutional encoder stated above. After that, we would have four encoded bit streams, denoted by s_1', s_2', s_3', s_4'. Notice that the encoded streams are not of the length of the raw data streams, for the rate-1/2 encoder used in our study here, the length of the encoded streams is doubled, we sacrifice the data rate by half to trade for better BER performance.

The four encoded bit streams are then being passed through the STBC mapper as before and sent via the MIMO system.

5.2.1 Hard-Decision Decoding

We get the estimated bit streams $\hat{s_1}'$, $\hat{s_2}'$, $\hat{s_3}'$, $\hat{s_4}'$ at the output of the hard-decision symbol detector, we then pass them through the Viterbi decoder to decode them back into the raw data streams. BER performance of such a hard-decision decoding system is shown comparing to the un-encoded system.

5.2.2 Soft-Decision Decoding

For a soft-decision decoder, the input is no longer a precise 0 or 1 binary estimation but the likelihood ratio $P(c=1)/P(c=0)$ of each bits.

To calculate the likelihood ratio of each individual bit we have to consider the receiver structure. For the differential encoding and decoding system in section 4.2, to decode the differentially encoded information message, the receiver

has to consider two consecutive received data blocks

$$\underbrace{\begin{bmatrix} \boldsymbol{Y}(t-1) & \boldsymbol{Y}(t) \end{bmatrix}}_{\boldsymbol{Y}_e} = \underbrace{\boldsymbol{H}\boldsymbol{X}(t-1)}_{\text{effective channel } \boldsymbol{H}_e} \cdot \underbrace{\begin{bmatrix} \boldsymbol{I} & \boldsymbol{U}(t) \end{bmatrix}}_{\boldsymbol{U}_e} + \underbrace{\begin{bmatrix} \boldsymbol{E}(t-1) & \boldsymbol{E}(t) \end{bmatrix}}_{\boldsymbol{E}_e} \tag{5.2}$$

where $\boldsymbol{Y}_e$ and $\boldsymbol{U}_e$ stand for the "effective" received data block and the transmitted matrix, respectively. To detect the transmitted matrix $\boldsymbol{U}_e$ given $\boldsymbol{Y}_e$, the noncoherent detection technique discussed in section 4.2 can be used. Though there is one main difference: For hard-decision decoding in section 4.2 we do not need to know the exact value of the likelihood function, but simply choose the $\boldsymbol{U}_e$ to maximize it. For this purpose we only need to concentrate the function against the unknown channel, and simply treat the noise variance σ^2 as an unknown constant would be good enough.

However for soft-decision decoding we have to not only maximize the likelihood function against $\boldsymbol{U}_e$, but have to evaluate the exact values of the likelihood function for any possible transmitted matrix $\boldsymbol{U}_e$, to achieve this we have to concentrate the likelihood function against both the unknown channel $\boldsymbol{H}$ and the unknown noise variance σ^2. The logarithm of the likelihood function is the following [16]

$$L(\boldsymbol{Y}_e|\boldsymbol{U}_e, \boldsymbol{H}_e, \sigma^2) = -2Nn_r \log \sigma^2 - \frac{1}{\sigma^2}\mathrm{Tr}\left\{(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)^H\right\} \tag{5.3}$$

Maximization of (5.3) with respect to σ^2 requires that

$$\begin{aligned} &\frac{\partial L(\boldsymbol{Y}_e|\boldsymbol{U}_e, \boldsymbol{H}_e, \sigma^2)}{\partial \sigma^2} = 0 \\ &\Rightarrow -2Nn_r\frac{1}{\sigma^2} + \frac{1}{\sigma^4}\mathrm{Tr}\left\{(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)^H\right\} = 0 \end{aligned} \tag{5.4}$$

which yields a positive solution for σ^2

$$\hat{\sigma}^2 = \frac{1}{2Nn_r}\mathrm{Tr}\left\{(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)^H\right\} \tag{5.5}$$

Inserting (5.5) into (5.3) gives

$$L_1(\boldsymbol{Y}_e|\boldsymbol{U}_e,\boldsymbol{H}) = -\log \mathrm{Tr}\left\{(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)(\boldsymbol{Y}_e - \boldsymbol{H}_e\boldsymbol{U}_e)^H\right\} \tag{5.6}$$

We next continue to concentrate the likelihood function to eliminate $\boldsymbol{H}_e$. Maximizing (5.6) with respect to $\boldsymbol{H}_e$, we get

$$\hat{\boldsymbol{H}}_e = \boldsymbol{Y}_e\boldsymbol{U}_e^H(\boldsymbol{U}_e\boldsymbol{U}_e^H)^{-1} = \frac{1}{2}\boldsymbol{Y}_e\boldsymbol{U}_e^H \tag{5.7}$$

where we used the fact that $\boldsymbol{U}_e\boldsymbol{U}_e^H = 2\boldsymbol{I}$.

Insertion of (5.7) into (5.6) gives the concentrated likelihood function

$$\begin{aligned}\hat{L}(\boldsymbol{Y}_e|\boldsymbol{U}_e) &= -\log \mathrm{Tr}\left\{(\boldsymbol{Y}_e - \frac{1}{2}\boldsymbol{Y}_e\boldsymbol{U}_e^H\boldsymbol{U}_e)(\boldsymbol{Y}_e - \frac{1}{2}\boldsymbol{Y}_e\boldsymbol{U}_e^H\boldsymbol{U}_e)^H\right\} \\ &= -\log \mathrm{Tr}\left\{\boldsymbol{Y}_e\boldsymbol{Y}_e^H - \frac{1}{2}\boldsymbol{Y}_e\boldsymbol{U}_e^H\boldsymbol{U}_e\boldsymbol{Y}_e^H\right\}\end{aligned} \tag{5.8}$$

For a particular information bit, say s_1, the likelihood ratio can then be approximated by

$$\begin{aligned}\frac{P(s_1=1)}{P(s_1=0)} &= \frac{\sum_{\boldsymbol{U}_e:s_1=1}\exp\left(L(\boldsymbol{Y}_e|\boldsymbol{U}_e)\right)}{\sum_{\boldsymbol{U}_e:s_1=0}\exp\left(L(\boldsymbol{Y}_e|\boldsymbol{U}_e)\right)} \\ &= \frac{\sum_{\boldsymbol{U}_e:s_1=1}\frac{1}{\mathrm{Tr}\left\{\boldsymbol{Y}_e\boldsymbol{Y}_e^H - \frac{1}{2}\boldsymbol{Y}_e\boldsymbol{U}_e^H\boldsymbol{U}_e\boldsymbol{Y}_e^H\right\}}}{\sum_{\boldsymbol{U}_e:s_1=0}\frac{1}{\mathrm{Tr}\left\{\boldsymbol{Y}_e\boldsymbol{Y}_e^H - \frac{1}{2}\boldsymbol{Y}_e\boldsymbol{U}_e^H\boldsymbol{U}_e\boldsymbol{Y}_e^H\right\}}}\end{aligned} \tag{5.9}$$

We can thus evaluate this likelihood ratio for each s_1, s_2, s_3, s_4 bits for every time intervals and use this soft information bit estimation as the input of the soft-decision Viterbi decoder.

Alternatively, the receiver may use the previously received data blocks to estimate the noise variance. Doing so we just need to insert the estimated noise variance into the likelihood function, and it is not necessary to concentrate the likelihood function with respect to the noise variance. If we estimate the noise

variance for a long enough time and take the average value, we will hopefully get a very accurate estimation. To get the estimated $\hat{\sigma}^2$ from the received data blocks we simply need to refer back to (5.5)

$$\hat{\sigma}^2 = \frac{1}{2Nn_r}\mathrm{Tr}\left\{(\boldsymbol{Y}_e - \hat{\boldsymbol{H}}_e\hat{\boldsymbol{U}}_e)(\boldsymbol{Y}_e - \hat{\boldsymbol{H}}_e\hat{\boldsymbol{U}}_e)^H\right\} \tag{5.10}$$

Substituting (5.7) into (5.10) gives

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{2Nn_r}\mathrm{Tr}\left\{(\boldsymbol{Y}_e - \frac{1}{2}\boldsymbol{Y}_e\hat{\boldsymbol{U}}_e^H\hat{\boldsymbol{U}}_e)(\boldsymbol{Y}_e - \frac{1}{2}\boldsymbol{Y}_e\hat{\boldsymbol{U}}_e^H\hat{\boldsymbol{U}}_e)^H\right\} \\ &= \frac{1}{2Nn_r}\mathrm{Tr}\left\{\boldsymbol{Y}_e\boldsymbol{Y}_e^H - \frac{1}{2}\boldsymbol{Y}_e\hat{\boldsymbol{U}}_e^H\hat{\boldsymbol{U}}_e\boldsymbol{Y}_e^H\right\}\end{aligned} \tag{5.11}$$

Substituting the $\hat{\sigma}^2$ and $\hat{\boldsymbol{H}}_e = \frac{1}{2}\boldsymbol{Y}_e\hat{\boldsymbol{U}}_e^H$ back into (5.3) yields the likelihood function with estimated noise variance

$$L(\boldsymbol{Y}_e|\boldsymbol{U}_e, \hat{\sigma}^2) = -2Nn_r\log\hat{\sigma}^2 - \frac{1}{\hat{\sigma}^2}\mathrm{Tr}\left\{\boldsymbol{Y}_e\boldsymbol{Y}_e^H - \frac{1}{2}\boldsymbol{Y}_e\hat{\boldsymbol{U}}_e^H\hat{\boldsymbol{U}}_e\boldsymbol{Y}_e^H\right\} \tag{5.12}$$

With an accurate $\hat{\sigma}^2$ we may expect that this likelihood function performs better than the concentrated function with unknown noise variance since there is one less unknown variable. Figure 5–2 shows the simulation results of the performance of both hard-decision and soft-decision schemes.

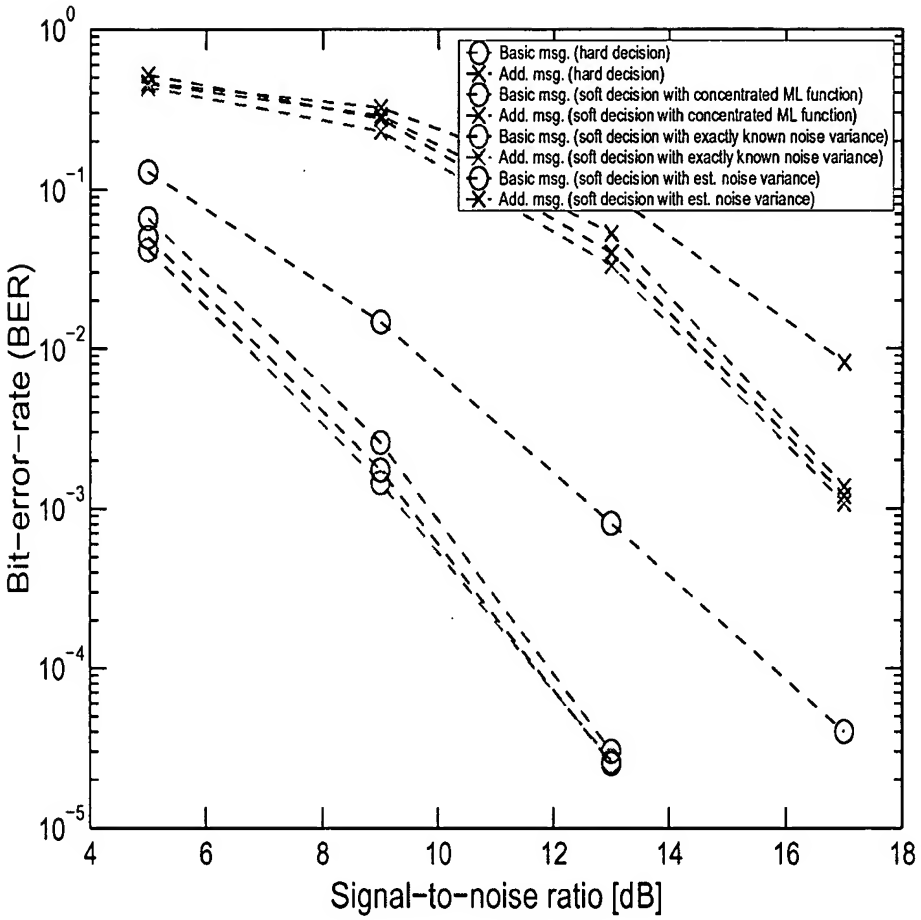

Figure 5–2: With rate-1/2 convolutional coding, comparison of BER performance are shown for hard-decision detection, soft-decision detection with concentrated likelihood function, and soft-decision with estimated noise variance.

CHAPTER 6
A NETWORK APPLICATION EXAMPLE

We consider employing our nonuniform space-time codes in a broadcasting telecommunication system. First we present a short introduction to contemporary typical wireless cellular systems.

The concept of a wireless cellular system is fairly simple: a large region to be served are divided into many small areas called "*cells*". Each cell is serviced by one or more base station(s) located at the center or the corners of the cell. The advantage of such a cellular system is that the available frequencies band can be reused in cells which are far apart; thus the same available frequency band can now be enjoyed by a small area which may consist of several small cells.

Figure 6–1 shows a 7-cell frequency reuse scheme. A cluster is formed by seven cells. A cluster can use all the available bandwidth, and the whole frequency band is being reused in all other clusters. So the whole frequency band is now shared by seven cells only, thus each user in the cluster should be able to get a adequate amount of usable bandwidth. Since the adjacent clusters use exactly the same frequency band, a power control scheme can be employed to keep the Signal-to-Interference-Ratio (SIR) under certain threshold.

Now we are moving on to see what is happening on one particular cell. Figure 6–2 shows a typical cell. For simplicity we assume the shape of the cell to be circular, and that the base station is located at the center. Let R_0 represents the cell radius. Typically the SNR at the cell border should be the lowest. To predict the signal strength at various places of the cell, some sort of propagation model has to be used. Here we first introduce the free space propagation model

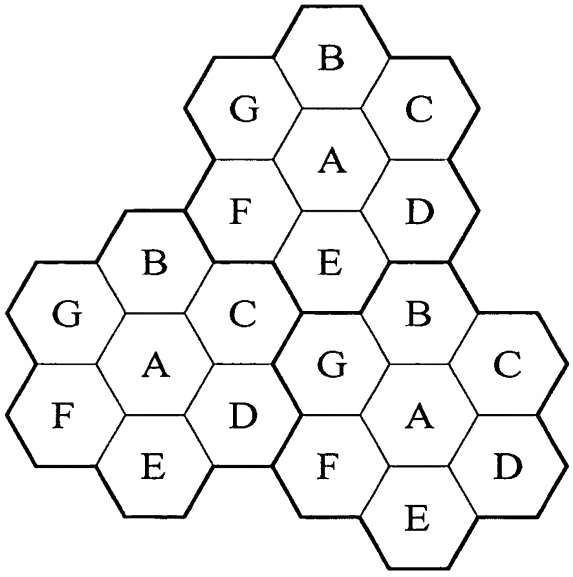

Figure 6–1: 7-cell frequency reuse system.

due to its simplicity. Assuming there exists a Line-Of-Sight (LOS) propagation path, then the received signal strength can be predicted by

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} \tag{6.1}$$

where P_t is the transmit power at the transmit antenna, G_t and G_r are the transmit and receive antenna gains respectively, λ is the wavelength of the transmitted signal, d is the propagation distance. In the previous discussion of space-time coding it is assumed that we have a frequency-flat channel, this means that we must use a narrowband modulated signal.

It is well known that the received signal strength attenuates as d^{-2} in free space theoretically, but in practical environment for a wireless link, the LOS path usually not exists and the propagation may encounter various loss such as scattering and other obstructions, so the attenuation is usually much higher. Even with a LOS path, the received signal strength usually attenuates faster than d^{-2} due to destructive reflection. Another commonly used model is the

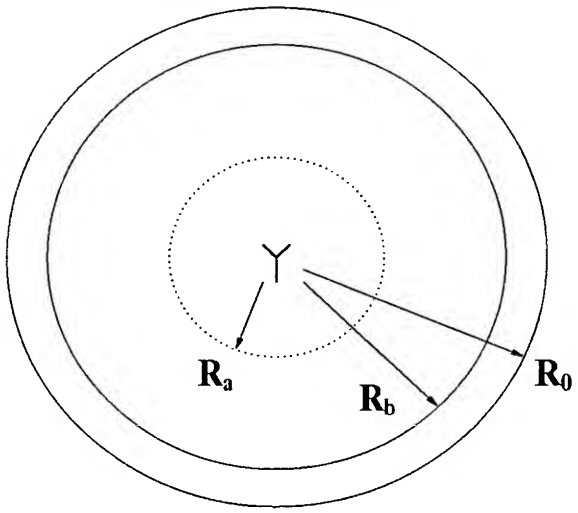

Figure 6–2: Typical cell layout for a wireless telecommunication system.

log-distance path loss model

$$P_r(d) = P_0 \cdot \left(\frac{d}{d_0}\right)^{-n} \tag{6.2}$$

Here P_0 is the received power at a reference point, d_0 is the distance between the transmitter and the reference point, and n is the path loss exponent. For ideal free space, $n = 2$. For an urban area we may take $n \approx 3$ or 4 [24].

Due to the unpredictability of the radio environment the above model can only represent the *average* received power. In fact the actual received power fluctuates vastly around the average value since the propagation environment varies at different time/space (varying reflective objects, obstructions, weather, etc). From a system design point of view the exact level of received power is quite hard to predict and this sort of power fluctuation may be modelled as a random quantity in a mathematical framework. So the *Log-normal Fading* model is being introduced

$$10\log_{10} P_r(d)\ [dB] = 10\log_{10} P_0\ [dB] - 10n\ log(\frac{d}{d_0}) + X_\sigma \tag{6.3}$$

where X_σ is a zero-mean Gaussian distributed random variable with variance σ^2 (Typical $\approx 6\ dB$ over large/small scale fading).

Thus we may say that for any point at the cell, the instantaneous SNR is a Gaussian random variable with the mean value given by the Log Distance Path Loss Model.

$$SNR = \overline{SNR} + X_\sigma \tag{6.4}$$

Since now the SNR at any point is modelled as a random variable then the instantaneous BER performance is also random, we can define the coverage area of a cell.

Define the coverage radius to be the radius of the area such that a specified detection performance (in terms of BER) is achievable with a certain probability that we call the Quality of Service (QoS). For example, if we let R_0 denote the 99% coverage radius then within the area of radius R_0 the message can be detected at a specified BER during 99% the of total time.

If nonuniform space-time codes are used to increase the data rate for the more capable receivers, it is clear that the coverage radius for the transmission of the basic message under the same QoS will decrease to $R_b < R_0$. Meanwhile the coverage radius of the transmission of the additional message now becomes R_a; of course its value depends on QoS for the additional message. For the simplicity of our analysis here let us assume the same QoS for both the basic and the additional message.

Denote the "tolerable" performance loss for the detection of the *basic* message as Δ dB. Assuming log-distance fading it can be shown that

$$\frac{R_b^2}{R_0^2} = 10^{-\frac{\Delta}{10}\cdot\frac{2}{n}} \tag{6.5}$$

$$\frac{R_a^2}{R_0^2} = 10^{-\frac{\Gamma}{10}\cdot\frac{2}{n}} \tag{6.6}$$

where Γ is the performance difference (in dB) between the detection of the *additional* message and that of the standard uniform constellation system. The parameter n is the path loss exponent for the radio link environment (i.e. Power $\propto$ (distance)$^{-n}$; for details see Rappaport et al. [24]). The above equations are also valid in a log-normal fading model. It is not difficult to obtain the result above: Denote the SNR required to have a certain BER performance as $SNR_{\text{th}}[dB]$. Then, for example, the 99% coverage radius R_0 can be determined as it satisfies

$$P\left[\overline{P_r(R_0)}[dB] + X_{sigma} > SNR_{\text{th}}[dB]\right] = 99\% \tag{6.7}$$

When we apply the nonuniform modulation the BER performance of the basic message deteriorates by Δ dB, which means that it requires a Δ dB higher SNR to obtain the same performance, so

$$P\left[\overline{P_r(R_b)}[dB] + X_{sigma} > SNR_{\text{th}}[dB] + \Delta\right] = 99\% \tag{6.8}$$

Comparing (6.7) and (6.8) we get

$$\begin{aligned} &\overline{P_r(R_b)}[dB] = \overline{P_r(R_0)}[dB] + \Delta \\ \Longrightarrow &10\log_{10} P_0\ [dB] - 10n\ log(\frac{R_b}{R_0}) = 10\log_{10} P_0\ [dB] + \Delta \end{aligned} \tag{6.9}$$

and (6.5) follows immediately, (6.6) can be shown similarly.

Figure 6–3 shows how R_b and R_a vary with Δ for the $R_b = 1$, $R_a = 1$ system, with the power attenuation factor of the radio link taken to be $n = 2$ and $n = 4$; these are typical values for free-space and downtown areas, respectively [24]. It can be seen that employing the nonuniform constellation does *not* increase the total throughput of the system (i.e., the sum of the data rate for all users) in the case $n = 2$. For example, if the coverage area for the basic message

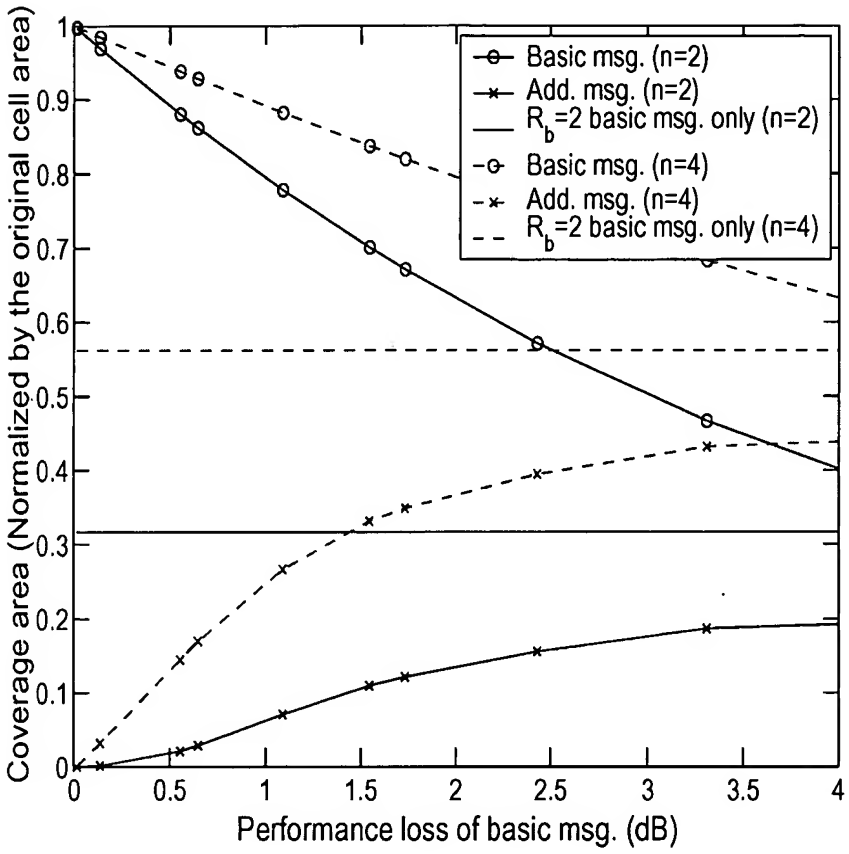

Figure 6–3: Coverage area for the basic/additional messages as a function of the "tolerable" performance loss, for the $R_b = 1$, $R_a = 1$ code. The horizontal curves represent the relative coverage area for the $R_b = 2$ code with a basic message only.

is allowed to decrease to 0.8 times of that of the original cell area, the coverage area of the additional message can only reach 0.15 times of that of the original cell area, which results in a decrease in overall system throughput. However for the case $n = 4$, the use of a nonuniform code increases the total throughput. This observation is not hard to understand; for $n = 4$ the area closely surrounding the base station enjoys a relatively much higher SNR than the area close to the border of the cell and hence the coverage area for the additional can be extended further. Note that the coverage areas of the nonuniform layered code (4.38) and

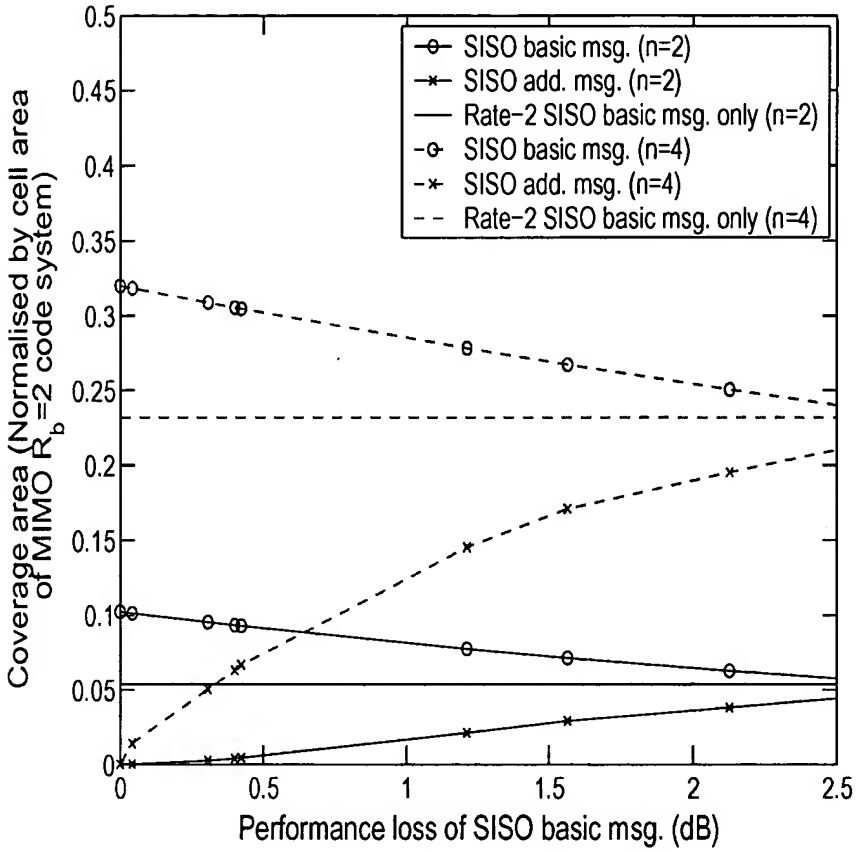

Figure 6–4: Coverage area for the SISO basic/additional messages as a function of the "tolerable" performance loss. The horizontal curves represent the relative coverage area for the SISO rate-2 code with a basic message only.

(4.39) (with $R_b = R_a = 1$) tend to approach that of the $R_b = 2$ code (4.46) when Δ increases.

The performance advantage of a 2-TX system upon 1-TX system is illustrated in Figure 6–4. The coverage area of the 1-TX system is only about 0.2 times as that of a 2-TX system for $n = 4$.

CHAPTER 7
CONCLUDING REMARKS AND FUTURE WORKS

We have presented new nonuniform space-time codes that can be encoded and detected differentially, and that are based entirely on phase-shift keying. We also discussed analytical criteria for code construction and optimization, and we compared its performance with that of a scheme based on the Alamouti code. We also studied a suboptimal detector and its performance, which we found to be satisfactory (a loss within 0.3 dB compared to joint ML decoding). We also demonstrated how nonuniform space-time codes in a broadcasting telecommunication system can increase the total throughput.

It may be argued that using already established nonuniform constellations for single transmit antenna systems together with, for instance, known linear space-time codes should be a natural approach to the problem of designing nonuniform space-time constellations. However, there are at least two problems associated with such an approach. First, it may in general *not* be optimal, simply because we are optimizing over the class of space-time codes and the class of nonuniform single-antenna constellations *separately*, instead of optimizing over the class of nonuniform MIMO constellations. Second, we found it difficult to incorporate constraints (such as constant envelope after differential encoding), that are desired or required from a practical implementation point of view. Therefore, we believe that it may be advantageous to design nonuniform space-time constellations by approaching the problem from first principles.

In this thesis, the nonuniform space-time block codes are chosen in quite an "arbitrary" way, and they are optimized with the help of chernoff bounds of

BER, which are always loose. We can expect the "optimal" nonuniform group codes suggested in this paper are indeed not optimal. Future works may include

1. Calculation of exact theoretical BER
2. Designing the nonuniform group codes in a "more" systematical way
3. Compare the performance of our nonuniform STBC to the theoretical capacity bound of the broadcast system

REFERENCES

[1] L-F Wei, "Coded modulation with unequal error probabilities," *IEEE Transactions on Communications*, vol. 41, pp. 1439–1449, Oct. 1993.

[2] A R Calderbank and N Seshadri, "Multilevel codes for unequal error protection," *IEEE Transactions on Information Theory*, vol. 39, pp. 1234–1248, July 1993.

[3] M Sajadieh, F R Kschischang, and A Leon-Garcia, "Modulation-assisted unequal error protection over the fading channel," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 900–908, Aug. 1998.

[4] K Ramchandran, A Ortega, K M Uz, and M Vetterli, "Multiresolution broadcast for digital HDTV using joint source/channel coding," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 6–23, 1993.

[5] W Li, "Overview of fine granularity scalability in the MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 301–317, Mar. 2001.

[6] M Memarzadeh, A Sabharwal, and B Aazhang, "Broadcast space-time coding," in *Proc. of the SPIE International Symposium on the Convergence of Information Theory and Communications (ITCOM)*, Denver, CO, Aug. 2001, pp. 112–120, SPIE, Bellingham, WA.

[7] C-H Kuo, C-S Kim, R Ku, and C-C J Kuo, "Embedded space-time coding for wireless broadcast," in *Proc. of Visual Communication and Image Processing*, San Jose, CA, Jan. 2002, pp. 189–197, SPIE, Bellingham, WA.

[8] M B Pursley and J M Shea, "Nonuniform phase-shift key modulation for multimedia multicast transmission in mobile wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 774–783, May 1999.

[9] T M Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, 1972.

[10] P P Bergmans and T M Cover, "Cooperative broadcasting," *IEEE Transactions on Information Theory*, vol. 20, pp. 317–324, May 1974.

[11] JPEG, "http://www.jpeg.org/," *Joint Photographic Experts Group*, 3/24/2003.

[12] M B Pursley and J M Shea, "Multimedia multicast wireless communications with phase-shift key modulation and convolutional coding," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1999–2010, 1999.

[13] M B Pursley and J M Shea, "Adaptive nonuniform phase-shift key modulation for multimedia traffic in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 1394–1407, 2000.

[14] E G Larsson and P Stoica, *Space-Time Block Coding for Wireless Communications*, Cambridge University Press, Cambridge, UK, 2003.

[15] Torsten Söderström and Petre Stoica, *System Identification*, Prentice Hall International, Hemel Hempstead, UK, 1989.

[16] E. G. Larsson, P Stoica, and J Li, "On maximum-likelihood detection and decoding for space-time coding systems," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 937–944, Apr. 2002.

[17] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[18] B. L. Hughes, "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[19] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, pp. 1169–1174, July 2000.

[20] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Transactions on Communications*, vol. 48, no. 12, pp. 2041–2052, Dec. 2000.

[21] G Ganesan and P Stoica, "Differential modulation using space-time block codes," *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 57–60, Feb. 2002.

[22] X G Xia, "Differential en/decoding orthogonal space-time block codes with APSK signals," *IEEE Communications Letters*, vol. 6, no. 4, pp. 150–152, Apr. 2002.

[23] V Tarokh, H Jafarkhani, and A R Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[24] T Rappaport, *Wireless Communications: Principles and Practice*, Prentice-Hall, Upper Saddle River, NJ, second edition, 2002.